On the dimension of the subfield subcodes of Hermitian codes

Ph. D. Thesis

by

Sabira El Khalfaoui

Thesis advisor:

Prof. Dr. Gábor P. Nagy

Doctoral School of Mathematics and Computer Science, Bolyai Institute
University of Szeged, Faculty of Science and Informatics

Szeged, 2020

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Gábor P. Nagy for his patience, motivation, and knowledge which supported my Ph.D. study and the related research. His guidance helped me to learn many things in different areas of research. His advice was invaluable for my research and the writing of this thesis.

My sincere thanks also go to both Bolyai Institute and Stipendium Hungaricum Foundation. They provided me an opportunity to join Ph.D. studies, and they gave access to all research facilities. Without their precious support, it would not be possible to conduct this research.

Large part of the results were obtained in the framework of the research project *Security Enhancing Technologies for the Internet of Things (SETIT)* [1]. The motivating discussions and feedback helped a lot for carrying out the presented work.

Last but not least, I would like to thank my family: my grandparents, my parents, my siblings, and my friends for supporting me spiritually throughout writing this thesis and my life in general.

[1]Project no. 2018-1.2.1-NKP-2018-00004 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2018-1.2.1-NKP funding scheme.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation seeks to study some classes of subfield subcodes of Hermitian codes. These papers [EKN19; EKN20] present the results of our research that aims at revealing the properties and the structure of the underlying classes of codes. Furthermore, we intend to examine the potential of this class of codes to improve the practicality of the McEliece cryptosystem. The problems we treat belong to coding theory and their applications to cryptography. They have a common aspect, which is security that refers to code-based cryptography. Here, we briefly introduce the preliminaries and the topics with a short history that describes the main results.

The result of the paper [EKN19] is discussed in chapter 3 which is about the proof of the true dimension of Hermitian subfield subcodes for specific parameters. Finding the true dimension of the subfield subcodes of linear codes was studied by many researchers who tried to improve the general bound of the dimension to obtain a code with a large dimension and minimum distance. We only present this problem for the class of Goppa codes in chapter 2. The solution to this problem allows us to find out more facts about the class of codes and which can later lead to further research.

In chapter 4, we rely on the paper [EKN20] which deals with the problem of approximat-

ing the true dimension of subfield subcodes of Hermitian codes by an explicit formula. We describe the statistical set up to tackle the experimental study to analyze the datasets of the true dimension of different subfield subcodes of Hermitian codes. The datasets were computed using our GAP package `HERmitian` [NEK19]. Based on adjusting the distribution to the underlying datasets using the method `fitmethis` of MATLAB [TM19; Cas20], we found that the extreme value distribution is the most suitable one.

Chapter 5 is dedicated to applying subfield subcodes of Hermitian codes in cryptography in which we precisely suggest the mentioned class of codes for McEliece cryptosystem. Mainly, we give a formula of the public key size in terms of the code rate using the result of the paper [EKN20], see also chapter 4. We describe an overview of post-quantum cryptography in which code-based cryptography is part of, representing the central area of applications concerning coding theory. This overview shows the importance of designing cryptographic schemes that can resist post-quantum attacks since the presence of quantum computer threatens the so-called classical cryptography. All cryptosystems are based on a computationally hard problem such as integer factorization (RSA), or discrete logarithm problem (ECC, ElGamal).

Chapter 2

Preliminaries

2.1 Error-correcting codes

In the last decades, there has been a huge need for reliable digital data transmission and storage systems. This need has grown thanks to the appearance of high-speed data networks for the interchange, the treatment and the storage of digital information in both the public and private sectors. It is necessary to incorporate communications and computer technology to design such systems. Obtaining a reliable data reproduction can be done by controlling the occurred errors which is the major aims of a designer.

In 1948, Shannon introduced a mathematical framework to describe communication channels with or without errors. In his famous paper [LC01], Shannon demonstrated the existence of encoding and decoding schemes. This work was to some extent inspired by Ludwig Boltzmann's work in statistical physics. Hamming gave the idea of detecting and correcting errors. It was a consequence of resolving the problem when his computer came to turn off every time it detected an error. Shannon's Second Theorem concerns channel coding. In other words, it adds extra information to a message that is intended to be sent in a noisy channel which protects it against transmission errors. Moreover, this

extra information permits us to detect or even correct some transmission errors which gave birth to error-correcting codes theory.

2.1.1 Block codes and Hamming distance

We assume that A is the set of q symbols which is called the alphabet. We denote A^n which is the set n-tuples $x = (x_1, \cdots, x_n)$, with $x_i \in A$.

Definition 2.1 (Block code)**.** *A block code C of length n over A is a nonempty subset of A^n. The elements of C are blocks called codewords. If C contains M codewords, then C has size M, and we denote it by an (n, M) code. If $M = q^k$, then C is called an $[n, k]$ code.*

The value $n - \log_q(M)$ is the redundancy of the code, which is the average number of symbols added to embed a message of size less than n into an n-tuple. We define the information rate as $R = \log_q(M)/n$.

Definition 2.2 (Encoding)**.** *An encoding of an $[n, k]$ block code C over A is a one-to-one map*

$$Enc: \quad A^k \quad \mapsto \quad A^n$$

where C is the image of A^k by Enc.

In order to evaluate the error-correcting capability of the code, we need a metric on A^n to measure the difference between two distinct words. The practical metric used in error-correcting codes is the *Hamming distance*.

Definition 2.3 (Hamming distance)**.** *Let $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_n)$ be two elements in A^n. The Hamming distance $d(x, y)$ is defined as the number of positions where x and y differ:*

$$d(x, y) = |\{i | x_i \neq y_i\}|.$$

The Hamming weight of a codeword x is $wt(x) = |\{i \,|\, x_i \neq 0\}|$.

Theorem 2.1. *The Hamming distance is a well-defined metric on A^n. It satisfies the following properties x, y, $z \in A^n$:*

- *Non-negativity: $d(x,y) \geq 0$, and $d(x,y) = 0 \iff x = y$.*
- *Symmetry $d(x,y) = d(y,x)$.*
- *Triangle inequality $d(x,z) \leq d(x,y) + d(y,z)$.*

Definition 2.4. *Let C be a non-empty subset of A^n. The minimum distance of C is defined as:*

$$d = d(C) = min\left\{d(x,y) | x, y \in C, x \neq y\right\}.$$

The main goal in the mathematical theory of error-correcting codes is to design codes of a given length and size for the largest possible minimum distance. This allows us to detect and correct a large number of errors. Moreover, useful codes can be characterized by efficient encoding and decoding algorithms.

2.1.2 Sphere packing bound

Definition 2.5.

- *Let $x \in A^n$, the ball of radius ρ centred at x is defined by*

$$B_\rho(x) = \{y \in A^n | d(x,y) \leq \rho\}$$

- *The sphere of radius ρ around x is defined by:*

$$S_\rho(x) = \left\{y \in A^n | d(x,y) = \rho\right\}.$$

Theorem 2.2. *Let x be an element of A^n, where A is an alphabet of q elements. Then*

$$|S_i(x)| = \binom{n}{i}(q-1)^i \text{ and } |B_r(x)| = \sum_{i=0}^{r}\binom{n}{i}(q-1)^i.$$

Theorem 2.3 (Sphere packing bound)**.** *Let $L_q(n,d)$ be the maximum number of codewords in a code C of length n over A and minimum distance at least d. The sphere packing bound is the following:*

$$L_q(n,d) \leq \frac{q^n}{\sum_{i=0}^{t} \binom{n}{i}(q-1)^i}, \quad t = \lfloor \frac{d-1}{2} \rfloor.$$

2.2 Linear codes

Linear codes are defined over finite fields. Let q be a prime power and let $A = \mathbb{F}_q$ be a finite field with q elements. In this case, $A^n = \mathbb{F}_q^n$ is a vector space. In particular, linear codes are more used than arbitrary ones since they have a structure, and they can be represented as the null space or the image of a linear transformation.

Definition 2.6. *A linear code C is a subspace of $\mathbb{F}_q^n$. It is denoted by $[n,k,d]_q$ or $[n,k,d]$, where n is the length, k is the dimension, and d is the minimum distance.*

It is clear that a linear $[n,k]$ code has q^k elements. The information rate is $R = k/n$ and the redundancy is $n-k$. In this case, the encoding is one-to-one linear transformation:

$$Enc: \ \mathbb{F}_q^k \ \mapsto \ \mathbb{F}_q^n$$

where $\mathbb{F}_q^k$ is the message space, and $Enc(\mathbb{F}_q^k) = C$. Enc can be represented by a $k \times n$ matrix that is called *generator matrix* denoted by G. This process consists of adding some redundancy to a message to produce a codeword. Notice that the rows of G form a basis for the linear code C.

A linear code C can be also defined by a null space of a matrix, which is called *parity check matrix* H in such a way that

$$C = \left\{ x \in \mathbb{F}_q^n \mid xH^\top = 0 \right\}.$$

The error-detection and error-correction capability of a linear code C can be determined by the mean of its minimum distance. This latest parameter controls the process of

decoding which consists of recovering the original message.

Theorem 2.4 (Singleton bound)**.** *Let C be an $[n, k]$ linear code. Then*

$$d(C) \leq n - k + 1.$$

A code that meets the Singleton bound is called maximum separable (MDS) code.

Definition 2.7 (Dual code)**.** *Let C be a $[n, k]$ linear code. The dual code $C^{\perp}$ of C is defined as follows:*

$$C^{\perp} = \left\{ x \in \mathbb{F}_q^n \mid c.x = 0, \text{ for all } c \in C \right\}.$$

$C^{\perp}$ is an $[n, n-k]$ linear code with generator matrix H, that is parity check matrix of C.

2.2.1 Subfield subcodes of linear codes

Definition 2.8. *Let C be a $[n, k]$ linear code over $\mathbb{F}_q$, where $q = r^m$ is a prime power. The $\mathbb{F}_q/\mathbb{F}_r$ subfield subcode $C|_{\mathbb{F}_r}$ of C is by definition the set*

$$C|_{\mathbb{F}_r} = C \cap \mathbb{F}_r^n$$

of all codewords in C with components in $\mathbb{F}_r$.

The $\mathbb{F}_q/\mathbb{F}_r$ subfield subcode is a linear (n, k_0, d_0) code with $d \leq d_0 \leq n$ and $n - k \leq n - k_0 \leq m(n-k)$. A parity check matrix of C over $\mathbb{F}_q$ yields at most $m(n-k)$ linearly independent parity equations over $\mathbb{F}_r$ for the subfield subcodes $C|_{\mathbb{F}_r}$.

In general, the minimum distance of the subfield subcode is bigger than the minimum distance of the original one.

Let $T_{\mathbb{F}_q/\mathbb{F}_r}$ be the trace polynomial in the field $\mathbb{F}_q$ with respect to $\mathbb{F}_r$, that is

$$T_{\mathbb{F}_q/\mathbb{F}_r}(x) = x + x^r + ... + x^{r^{m-1}}.$$

For a vector $c \in \mathbb{F}_q^n$, $T_{\mathbb{F}_q/\mathbb{F}_r}(c) = (T_{\mathbb{F}_q/\mathbb{F}_r}(c_1), \cdots, T_{\mathbb{F}_q/\mathbb{F}_r}(n))$. For a linear code C of length n and dimension k over $\mathbb{F}_q$, $T_{\mathbb{F}_q/\mathbb{F}_r}(C)$ is a linear code with the same length of C and dimension k_1 over $\mathbb{F}_r$.

Delsarte has come up with a very important result which relates the subfield subcode to the trace code in the following theorem:

Theorem 2.5 ([Del75])**.** *Let C be a $[n, k]$ linear code over $\mathbb{F}_q$. Then $(C|_{\mathbb{F}_r})^\perp = T_{\mathbb{F}_q/\mathbb{F}_r}(C^\perp)$ holds.*

The class of subfield subcodes and trace codes held the attention of many researchers. A lot of work was done on the class of subfield subcodes by Stichtenoth [Sti90], and it was improved upon in [SMS97]. The study of trace codes was made by Van der Vlugt [Vlu91; VDV91]. Roseiro stated the relation between trace codes and Goppa codes which was established in [Ros+92] using the tool given by Delsarte (see [Del75]).

Lemma 2.6. *Let C be an $[n, K]$ linear codes over the finite field $\mathbb{F}_q$, where $q = r^m$. The subfield subcode of C satisfies:*

$$\dim_{\mathbb{F}_r}(C \cap \mathbb{F}_r^n) = n - m(n - K) + \dim_{\mathbb{F}_r}(\ker(T_{\mathbb{F}_q/\mathbb{F}_r})). \tag{2.1}$$

Proof. In general, the dimension of the subfield subcode of C satisfies:

$$\dim_{\mathbb{F}_r}(C \cap \mathbb{F}_r^n) \leq K, \tag{2.2}$$

and the dimension of the trace code of C has the following bound:

$$\dim_{\mathbb{F}_r}(T_{\mathbb{F}_q/\mathbb{F}_r}(C)) \leq mK. \tag{2.3}$$

From both equations 2.2, 2.3, we have

$$\begin{aligned} n &= \dim_{\mathbb{F}_r}(C \cap \mathbb{F}_r^n) + \dim_{\mathbb{F}_r}(T_{\mathbb{F}_{q/\mathbb{F}_r}}(C^{\perp})) \\ &\leq \dim_{\mathbb{F}_r}(C \cap \mathbb{F}_r^n) + m(n-K). \end{aligned}$$

Which means that

$$\dim_{\mathbb{F}_r}(C \cap \mathbb{F}_r^n) \geqslant n - m(n-K). \tag{2.4}$$

Moreover, we have

$$\begin{aligned} m(n-K) &= \dim_{\mathbb{F}_r}(T_{\mathbb{F}_{q/\mathbb{F}_r}}(C^{\perp})) + \dim_{\mathbb{F}_r}(\ker(T_{\mathbb{F}_{q/\mathbb{F}_r}})) \\ &= n - \dim_{\mathbb{F}_r}(C \cap \mathbb{F}_r^n) + \dim_{\mathbb{F}_r}(\ker(T_{\mathbb{F}_{q/\mathbb{F}_r}})), \end{aligned}$$

Thus

$$\dim_{\mathbb{F}_r}(C \cap \mathbb{F}_r^n) = n - m(n-K) + \dim_{\mathbb{F}_r}(\ker(T_{\mathbb{F}_{q/\mathbb{F}_r}})). \tag{2.5}$$

□

This formula allows us to look for specific parameters of the subfield subcodes that can increase their dimension.

2.2.2 Reed-Solomon codes

In 1960, Irving Reed and Gustave Solomon constructed codes by evaluating polynomials on finite field elements. These codes were named after their inventors Reed-Solomon codes. In order to be useful in practice, it is not enough for a code to have a nice structure: an efficient decoding algorithm is needed as well.

In 1969, Berlekamp and Massey gave a decoding algorithm that made Reed-Solomon codes useful. The most notable applications of Reed-Solomon codes are in compact disks, bar codes and QR codes. From Reed-Solomon codes, we can obtain another class of codes that are defined over prime subfields and known as subfield subcodes of Generalized Reed-Solomon codes [Del75]. The class of subfield subcodes represents an area of great interest.

Definition 2.9. *Let q be a prime power, n an integers with $2 \leqslant n \leqslant q$. Let $\alpha = \{\alpha_1, ..., \alpha_n\} \subset \mathbb{F}_q$, $v = (v_1, ..., v_n)$ a nonzero vector of $\mathbb{F}_q$. The Generalized Reed-Solomon code, denoted by $\mathbf{GRS}_k(\alpha, v)$ consists of all vectors*

$$(v_1 F(\alpha_1), v_2 F(\alpha_2), ..., v_n F(\alpha_n)),$$

where $F(z) \in V_k = \{f(z) \in \mathbb{F}_q[z] \mid \deg(f) < k\}$.

Let $\left\{1, X, X^2, ..., X^{k-1}\right\}$ be one basis for the polynomial vector space V_k, then a generator matrix of $\mathbf{GRS}_k(\alpha, v)$ is

$$\begin{bmatrix} v_1 & \dots & v_n \\ \alpha_1 v_1 & \dots & \alpha_n v_n \\ \vdots & \ddots & \vdots \\ \alpha_1^{k-1} v_1 & \dots & \alpha_n^{k-1} v_n \end{bmatrix}$$

generalized Reed-Solomon codes are closed under duality. The dual of $\mathbf{GRS}_k(\alpha, v)$ is given by $\mathbf{GRS}_{n-k}(\alpha, v')$, where $v' \in \mathbb{F}_q^n \setminus \{0\}$.

Theorem 2.7. *The dual of $\mathbf{GRS}_k(\alpha, v)$ is $\mathbf{GRS}_{n-k}(\alpha, v')$ for some non zero $v' \in \mathbb{F}_q^n$. Moreover, v' depends on v but not on k.*

The definition of the usual Reed-Solomon code corresponds to the choice $v = (1, \dots, 1)$.

2.2.3 Goppa codes

The class of Goppa codes (introduced in 1970 [Gop70] by V.G. Goppa) contains good codes over $\mathbb{F}_q$ which asymptotically meet the Varshamov–Gilbert bound. In the general case, they are viewed as the subfield subcodes of Generalized Reed-Solomon codes. They form an important subclass of algebraic error-correcting codes.

Definition 2.10. *Let $g(z) \in \mathbb{F}_q[z]$, $L = \{\alpha_1, ..., \alpha_n\} \subset \mathbb{F}_q$ in such a way that $\forall i$, $g(i) \neq 0$. The Goppa Codes $\Gamma(L, g)$, of length n over $\mathbb{F}_r$, is the set of codewords, i.e. $n-$tuples*

$(c_1, ..., c_n) \in \mathbb{F}_r^n$, *satisfying*

$$\sum_{i=1}^{n} \frac{c_i}{z - \alpha_i} \equiv 0 \,(\mod g(z)) \,.$$

The simplest parity check matrix of $\Gamma(L, g)$ is

$$\begin{bmatrix} g(\alpha_1)^{-1} & \dots & g(\alpha_n)^{-1} \\ \alpha_1 g(\alpha_1)^{-1} & \dots & \alpha_n g(\alpha_n)^{-1} \\ \vdots & \vdots & \vdots \\ \alpha_1^{\deg g(z)-1} g(\alpha_1)^{-1} & \dots & \alpha_n^{\deg g(z)-1} g(\alpha_n)^{-1} \end{bmatrix}$$

Proposition 2.8. $\Gamma(L, g)$ *is the restriction to* $\mathbb{F}_r$ *of the dual of* $\mathbf{GRS}_{\deg g(z)}(\alpha, v)$

$$\Gamma(L, g) = \mathbf{GRS}_{n-\deg g(z)}(\alpha, v') \cap \mathbb{F}_r^n$$

with $v_i' = \frac{g(\alpha_i)}{\prod_{i \neq j}(\alpha_i + \alpha_j)}$.

Moreover, we consider $\mathbf{GRS}_{degg(z)}(L, v)$ code, then $T_{\mathbb{F}_q/\mathbb{F}_r}\left(\mathbf{GRS}_{\deg g(z)}(L, v)\right)$ equals the dual of Goppa code. Thus:

$$\Gamma\left(L, g\right)^{\perp} = T_{\mathbb{F}_q/\mathbb{F}_r}\left(\mathbf{GRS}_{\deg g(z)}(L, v)\right).$$

Applying Delsarte's result [Theorem 2.5,], we obtain a general formula for the dimension of any Goppa code [Vér05]:

$$k = n - m \deg g(z) + \dim_{\mathbb{F}_r} \ker\left(T_{\mathbb{F}_q/\mathbb{F}_r}\right). \tag{2.6}$$

Binary Goppa codes play a distinguished role in the theory. The reason for this is the fact that separable polynomials produce codes with twice better decoding threshold.

Theorem 2.9 ([MS77, Section 12.3, Theorem 6])**.** *The dimension of* $\Gamma(L, g)$ *and its minimal distance* d *satisfy*

- $k \geq n - m \deg g(z)$
- $d \geq \deg \bar{g}(z) + 1$

$\bar{g}(z)$ is the lowest degree perfect square which is divisible by $g(z)$. If $g(z)$ is separable polynomial of degree δ, then $d \geq 2\delta + 1$.

2.3 The true dimension of binary Goppa codes

By choosing the parameters of the binary Goppa code in an appropriate way, it is possible to increase its dimension and minimum distance. In this section, we investigate the problem of finding the true dimension of Goppa codes which is considered as subfield subcodes of generalized Reed-Solomon codes. We summarize two strategies that are used to get new bound for the dimension of Goppa codes which was the aim of many researchers [Ros+92; BS95; Véro01; Vér05; Vér98].

2.3.1 First strategy

The first strategy is about the link between the parity check matrix $\tilde{H}$ of Goppa codes and the parity check matrix H of **GRS** codes. The parity check matrix defined above does not generate the dual of Goppa codes because it is defined over $\mathbb{F}_q$. We can compute the parity check matrix $\tilde{H}$ over $\mathbb{F}_r$ from the parity check matrix H over $\mathbb{F}_q$ by converting each column vector of H to a column vector over $\mathbb{F}_r$. Therefore, computing the dimension of $\Gamma(L, g)$ is equivalent to computing the rank of $\tilde{H}$. H has $\deg g(z)$ rows, then $\tilde{H}$ has $m \deg g(z)$ rows which are not necessarily independent. This strategy has been stated in [Vér98], where the author explained (with an example [Vér05]) how to improve the bound $k \geq n - m \deg g(z)$, by looking for some polynomials and choosing a special basis, when computing $\tilde{H}$ from H, in order to find linear dependent rows [Vér05].

2.3.2 The trace Goppa codes

In [Vér98], Goppa codes are defined by the polynomial $g(z) = a(z)T_{\mathbb{F}_{r^{ms}}/\mathbb{F}_{r^s}}(b(z))$, where r is a prime number, m and s are two integers with $m > 1$, $a(z)$ and $b(z)$ are two arbitrary elements of $\mathbb{F}_{r^{ms}}[z]$. The authors proposed new bounds depending on m and s to reveal that the general bound cannot be achieved.

Definition 2.11. *Let $a(z)$ and $b(z)$ be two arbitrary elements of $\mathbb{F}_{r^{ms}}$. $\Gamma(L, g)$ is a Trace Goppa code iff*

$$g(z) = a(z)T_{\mathbb{F}_{r^{ms}}/\mathbb{F}_{r^s}}(b(z))$$

and

$$L = \mathbb{F}_{r^{ms}} \setminus \{z \in \mathbb{F}_{r^{ms}}, g(z) = 0\}.$$

This code will be denoted by $\Gamma_{r,m,s}(a(z), b(z))$.

Now, we describe Véron's bound on the dimension of the trace Goppa codes, with different choices of the polynomials $a(z)$ and $b(z)$. Each case has been studied with a specific form of the polynomials, the set L of $\mathbb{F}_{r^{ms}}$, and the basis of $\mathbb{F}_{r^{ms}}$ over $\mathbb{F}_r$ (for more details see [Vér98]). We present each case with its new bound as follows:

- **General Case :**

  the dimension k of $\Gamma_{r,m,s}(a(z), b(z))$ satisfies $k \geq n - ms \deg g(z) + (m-1)s$.

- **Quadratic Case:**

  the dimension k of $\Gamma_{r,2,s}(a(z), b(z))$, satisfies $k \geq n - 2s \deg g(z) + 2s - 1$.

- **Particular Binary Case:**

  the dimension k of $\Gamma_{2,m,s}(1, b(z))$, satisfies $k \geq n - ms \deg g(z) + ms$.

- **Binary-Quadratic Case :**

  the dimension k of $\Gamma_{2,2,s}(a(z), b(z))$, satisfies $k \geq n - 2s \deg g(z) + 3s - 1$.

2.3.3 Second Strategy

This strategy used Delsarte's result [Del75] so as to define codes with large dimension k. It is based on using the image of the dual code under the trace map with a rank that is equal to redundancy [Ros+92]. The objective of the strategy is to find polynomials $g(z)$ such that the trace map has a large kernel. This strategy was the main idea of [Ros+92], it was applied to the classes of primitive binary Goppa codes whose polynomial satisfies $G^{2^s}(X) \equiv G(X)(\mod X^{2^{2s}} + X)$.

Here, we describe some special cases for which the authors in [Ros+92] stated lower bounds for the dimension. We set $r = 2$ and $m = 2$. The authors consider primitive, separable, binary Goppa codes of length n and dimension k with locator field $L = GF(2^{2s})$. They took three Goppa polynomials, $G_1(X) = X^{2^s} + X$, $G_2(X) = X^{2^s+1} + 1$ and $G_3(X) = G_1(X)/H(X)$, where $H(x) = X$ or $H(X) = X + 1$. For each case we take $\pi(X)$ in such a way that $G_i(X)\pi(X) = X^{2^{2s}} + X$ for $i = 1, 2, 3$, so $n = 2^{2s} - r$ where $r = \deg(G_i(X))$.

The lower bound on the dimension for these polynomials are:

- $k \geq n - 2s \deg G_1(X) + 3s - 1$.
- $k \geq n - 2s \deg G_2(X) + 5s$.
- $k \geq n - 2s \deg G_3(X) + s - 1$.

In his dissertation, using a computer, Roseiro checked that these bounds are reached for $s = 2, 3, 4, 5$. The authors of [Ros+92] called an open problem to prove this for all $s \geq 2$. In [Vér05], it is an open problem to know if it was reached for $s \geq 5$.

In [Ros+92], the authors studied the dimension of binary Goppa codes with $g(z) = z^{2^s} + z$. They gave a new bound for the dimension:

$$\dim \Gamma(L, g) \geq n - 2s \deg g(z) + 3s - 1.$$

A generalization of this result has been described in [Vér98] where the trace Goppa codes has been introduced with the trace Goppa polynomial $g(z) = a(z)T_{\mathbb{F}_{r^{ms}}/\mathbb{F}_{r^s}}(b(z))$.
In [Vér98] for the binary quadratic case ($r = 2$, $m = 2$), the dimension of binary Goppa codes defined by $g(z) = a(z)T_{\mathbb{F}_{2^{2s}}/\mathbb{F}_{2^s}}(b(z))$, where $a(z) = 1$ and $b(z) = z$ satisfies

$$\dim \Gamma(L, g) \geq n - 2s \deg g(z) + 3s - 1.$$

In this case, the aim of [Véro01] is to prove one of the conjectures of [Ros+92], where the author showed that the dimension of Goppa codes achieved the lower bound in the following theorem:

Theorem 2.10 ([Véro01])**.** *Let $g(z) = T_{\mathbb{F}_{2^{2s}}/\mathbb{F}_{2^2}}(z)$ and $L = \mathbb{F}_{2^{2s}} \setminus \mathbb{F}_{2^s}$, the dimension of the Goppa code $\Gamma(L, g)$ satisfies*

$$\dim \Gamma(L, g) = n - 2s \deg g(z) + 3s - 1.$$

In [Vér05], Véron proved the two remaining conjectures of [Ros+92] on the true dimension of two classes of Goppa codes. The author started by proving the third conjecture, which is about defining the dimension of Goppa codes given by $g_3(z)$. As he mentioned, the proof is different from that of the first conjecture (which is proved in [Véro01]) and used the result in [BS95; Vér98]. He considered polynomials of the form $g_1(z)/(z + \beta)$ for any $\beta \in \mathbb{F}_{2^s}$.

2.3.4 True dimension of $\Gamma(L_3, g_3)$

Let $\beta \in \mathbb{F}_{2^s}$ and denote by $g_{3,\beta}$ the polynomial $(z^{2^s} + z)/(z + \beta)$, and let $L_{3,\beta}$ be the set $\{\beta, \alpha_1, ..., \alpha_n\} = L_1 = \mathbb{F}_{2^{2s}} \setminus \mathbb{F}_{2^s}$.

Theorem 2.11 ([Vér05])**.** *For all $\beta \in \mathbb{F}_{2^s}$, the dimension of the Goppa code $\Gamma(L_{3,\beta}, g_{3,\beta})$ is*

$$n - 2s \deg g_{3,\beta}(z) + s - 1.$$

To prove the main result, the author used the following lemmas:

Lemma 2.12. *All codes $\Gamma(L_{3,\beta}, g_{3,\beta})$ are equivalent to $\Gamma(L_{3,0}, g_{3,0})$.*

Lemma 2.13. *Let $c = (c_0, c_1, ..., c_n) \in \mathbb{F}_{2^{n+1}}$ and $\omega(c)$ be the Hamming weight of c, then*

$$c \in \Gamma(L_{3,0}, g_{3,0}) \text{ and } \omega(c) \text{ even } \Leftrightarrow c_0 = 0 \text{ and } c = (c_0, c_1, ..., c_n) \in \Gamma(L_1, g_1).$$

2.3.5 True dimension of $\Gamma(L_2, g_2)$

Theorem 2.14 ([Vér05])**.** *Let $g(z) = z^{2^s+1} + 1$, $L = \mathbb{F}_{2^{2s}} \setminus \{z \in \mathbb{F}_{2^{2s}} \mid g(z) = 0\}$ and $n = card(L)$, the dimension of the Goppa code $\Gamma(L, g)$ satisfies:*

$$k = n - 2s \deg g(z) + 5s.$$

We checked the results above for small small values of q by our implementation in GAP [Gap] using the GAP package `GZero` [Nag17].

Chapter 3

Algebraic geometry codes

In this chapter, we deal with codes constructed from geometric objects, which are usually called algebraic geometry codes over a finite field. These codes are built from algebraic curves. They can be defined by *evaluating functions* or by using *residues of differentials*. Many parameters and properties can be immediately derived from well-known theorems for algebraic curves. Constructing codes using the theory of algebraic geometry reveals a geometric description of these codes and their duals as well. Also, one can have good lower bounds for their parameters expressed in terms of invariants of algebraic curves. In the following, we describe the general idea of algebraic geometry codes using the functional approach.

Let $\mathcal{X}$ be a geometric object and $\mathcal{P}$ be a set of n points $\mathcal{P} = \{P_1, \cdots, P_n\}$. We assume that the set of function on $\mathcal{X}$ over a finite field $\mathbb{F}_q$ forms a vector space $\mathscr{L}$ over $\mathbb{F}_q$, such that for all i, $f(P_i) \in \mathbb{F}_q$ and $f \in \mathscr{L}$. Then one can obtain an evaluation map

$$\begin{array}{rccl} ev: & \mathscr{L} & \mapsto & \mathbb{F}_q^n \\ & f & \mapsto & (f(P_1), \cdots, f(P_n)), \end{array}$$

which is a linear map, and its image set represents a linear code.

The construction above is a generalization of the construction of Reed-Solomon codes. That is a typical example of AG codes, where the geometric object is a projective line over $\mathbb{F}_q$ (see section 3.2.1).

In the same way, we define another vector space Ω of differentials on $\mathcal{X}$. Then one takes residues of differentials to define another class of linear codes as the image of the following map

$$\begin{aligned} res: \ \Omega &\mapsto \mathbb{F}_q^n \\ \omega &\mapsto (res_{P_1}(\omega), \cdots, res_{P_n}(\omega)). \end{aligned}$$

This is also a linear map, its image set represents a linear code over $\mathbb{F}_q$.

3.1 Algebraic geometry overview

3.1.1 Algebraic curves, places, divisors

Let $f(X,Y) \in K[X,Y]$ be an absolutely irreducible polynomial of degree n (it is also irreducible over the algebraic closure $\overline{K}$). Let $\mathcal{X}$ be the associated algebraic plane curve over K denoted by $\mathcal{X} : f(X,Y) = 0$. The set of affine points is as follows:

$$\mathcal{X}(\overline{K}) = \left\{(x,y) \in \overline{K} \times \overline{K} \mid f(x,y) = 0\right\}.$$

We denote the set of rational points by $\mathcal{X}(K)$ that is defined as:

$$\mathcal{X}(K) = \{(x,y) \in K \times K \mid f(x,y) = 0\}.$$

A point $P = (x,y)$ of $\mathcal{X}$ is nonsingular or smooth if there exists a tangent line to the curve $\mathcal{X}$ at P, in other words, if $\left(\frac{\partial f}{\partial X}(x,y), \frac{\partial f}{\partial Y}(x,y)\right) \neq (0,0)$. $\mathcal{X}$ is called smooth (nonsingular) if every point $P \in \mathcal{X}$ is smooth, which implies that f is absolutely irreducible. The genus is the most important invariant of an algebraic curve. It is a non-negative integer, which

is given by the genus formula

$$g = \frac{(n-1)(n-2)}{2}$$

for smooth curves of degree n.

Now, we deal with projective curves rather than affine plane curves. The homogenous equation of the affine curve $\mathcal{X} : f(x, y) = 0$ is $F(X, Y, Z) = 0$ where $F(X, Y, Z) = Z^n f(X/Z, Y/Z)$. The projective points are defined as zeros of the homogenous polynomial $F(X, Y, Z)$. In particular, the affine point (x, y) of $\mathcal{X}$ is represented by the point $(x : y : 1)$ in projective coordinates. A projective point (x, y, z) of $\mathcal{X}$ is said to be at infinity when $z = 0$.

A divisor on $\mathcal{X}$ is a formal sum $D = n_1P_1 + \cdots + n_kP_k$ where the coefficients $n_1, \cdots, n_k$ are integers and $P_1, \cdots, P_k$ are points of $\mathcal{X}$. The degree of the divisor D is $\deg D = \sum_{i=0}^{k} n_i$. The valuation of the divisor D at a point P_i is $v_{P_i}(D) = n_i$. The support of D is the set $\{P_i \mid n_i \neq 0\}$.

3.1.2 Function fields and Riemann-Roch spaces

Let $\mathcal{X} : f(X, Y) = 0$ be a smooth plane algebraic curve. The function field $K(\mathcal{X})$ of $\mathcal{X}$ is generated by the elements x, y satisfying the algebraic relation $f(x, y) = 0$.

Let h be a non-zero function of $K(\mathcal{X})$. We denote by $\mathrm{Div}(h)$ the principal divisor. The degree of such a divisor is zero. We can define the divisor of h to be

$$\mathrm{Div}(h) = \mathrm{Div}(h)_0 - \mathrm{Div}(h)_\infty,$$

where $\mathrm{Div}(h)_0$ is said to be the divisor of zeros of h and $\mathrm{Div}(h)_\infty$ is the divisor of its poles.

Let K' be a vector space over $K(\mathcal{X})$. We denote by $D_{\mathcal{X},K'}$ the set of all derivations $D : K(\mathcal{X}) \mapsto K'$, and $D_{\mathcal{X}}$ when $K' = K(\mathcal{X})$. Furthermore, for every separable function $h \in K(\mathcal{X})$, dh is the exact differential arising from h. We denote the set of all differentials by Ω. Also, $\mathrm{res}_P(dh)$ is the residue of dh at a point P of $K(\mathcal{X})$.

For any divisor A of $K(\mathcal{X})$, the *Riemann-Roch space* of A is

$$\mathscr{L}(A) = \{h \in K(\mathcal{X}) \setminus \{0\} | \operatorname{Div}(h) \succeq -A\} \cup \{0\}.$$

We denote the dimension by $\ell(A) = \dim(\mathscr{L}(A))$. Furthermore, the *differential space* of A is

$$\Omega(A) = \{dh \in \Omega \mid \operatorname{Div}(dh) \succeq A\} \cup \{0\}.$$

The index of specialty for a divisor A is the integer

$$i(A) = \ell(A) - \deg A + g - 1.$$

Both the Riemann-Roch space and the differential space are linear spaces over K. Their dimensions are obtained by the theorem of Riemann-Roch:

Theorem 3.1 (Riemann-Roch)**.** *Let $\mathcal{X}$ be a smooth curve over K of genus g. Let A be an arbitrary divisor of the function field $K(\mathcal{X})$. Then we have*

$$\ell(A) = \deg(A) + 1 - g + \ell(W - A),$$

where W is a canonical divisor of $K(\mathcal{X})$.

An immediate corollary is the inequality

$$\ell(A) \geq \deg(A) + 1 - g.$$

for an arbitrary divisor A. Moreover, since W has degree $2g-2$,

$$\ell(A) = \deg(A) + 1 - g$$

provided $\deg(A) > 2g-2$.

3.1.3 Algebraic plane curves over finite fields

Let q be a prime power of p, and $\overline{\mathbb{F}}_q$ the algebraic closure of $\mathbb{F}_q$. We denote $Frob_q$ the Frobenius automorphism

$$\begin{array}{rccc} Frob_q: & \overline{\mathbb{F}}_q & \mapsto & \overline{\mathbb{F}}_q \\ & x & \mapsto & x^q. \end{array}$$

This map can be extended to $\mathbb{F}_q$–polynomials by coefficients and to affine and projective points over $\overline{\mathbb{F}}_q$ by coordinates.

Let $\mathcal{X}$ be a curve over $\mathbb{F}_q$, and P one of its points, then $Frob_q(P)$ is also a point of $\mathcal{X}$. $\mathcal{X}$ is $\mathbb{F}_q$–rational curve if it is invariant under the action of $Frob_q$. Similarly, $\mathbb{F}_q$–rational places and divisors are invariant under the action of the Frobenius automorphism.

3.2 Algebraic geometry codes (AG codes)

Let q be a prime power, and $\mathbb{F}_q$ be the finite field of order q. Let $\mathcal{X}$ be an algebraic curve, i.e., an affine or projective variety of dimension one, which is absolutely irreducible and nonsingular and whose defining equations are (homogeneous) polynomials with coefficients in $\mathbb{F}_q$. Let g be the genus of $\mathcal{X}$. In the following, $P_1, \cdots, P_n$ are pairwise distinct places on $\mathcal{X}$ and D is the divisor $D = P_1 + ... + P_n$. Furthermore, G is another divisor with support disjoint from D.

Definition 3.1. *The algebraic geometry code $C_{\mathscr{L}}(D, G)$ associated with the divisors D and G is defined as*

$$C_{\mathscr{L}}(D, G) = \{(f(P_1), f(P_2), ..., f(P_n)) \mid f \in \mathscr{L}(G)\} \subseteq \mathbb{F}_q^n.$$

In other words, $C_{\mathscr{L}}(D, G)$ is the image of $\mathscr{L}(G)$ under the evaluation map

$$\mathscr{L}(G) \ni f \mapsto (f(P_1), ..., f(P_n)) \in \mathbb{F}_q^n.$$

Theorem 3.2 ([Sti09])**.** *$C_{\mathscr{L}}(D, G)$ is a $[n, k, d]$ codes with parameters:*

- *$k = \ell(G) - \ell(G - D)$ where $\ell(G) = \dim \mathscr{L}(G)$*
- *$d \geq n - \deg G$*

Notice that the condition $n > \deg G$ implies the evaluation map $\mathscr{L}(G) \to \mathbb{F}^n$ to be injective. If $n \leq \deg G$, then it is possible that $C_{\mathscr{L}}(D, G)$ has dimension less than n and positive true minimum distance. However, this case cannot be described by the Riemann-Roch theorem.

Now, we define another class of AG codes called differential codes, which is an alternative of $C_{\mathscr{L}}(D, G)$ by taking the differential space $\Omega(G)$ rather than the Riemann-Roch space $\mathscr{L}(G)$. We will see later that it is useful to have both families of AG codes when tackling decoding algorithms.

Definition 3.2. *Let G and D be divisors as before. We define the code $C_\Omega(D, G) \subseteq \mathbb{F}_q$ by*

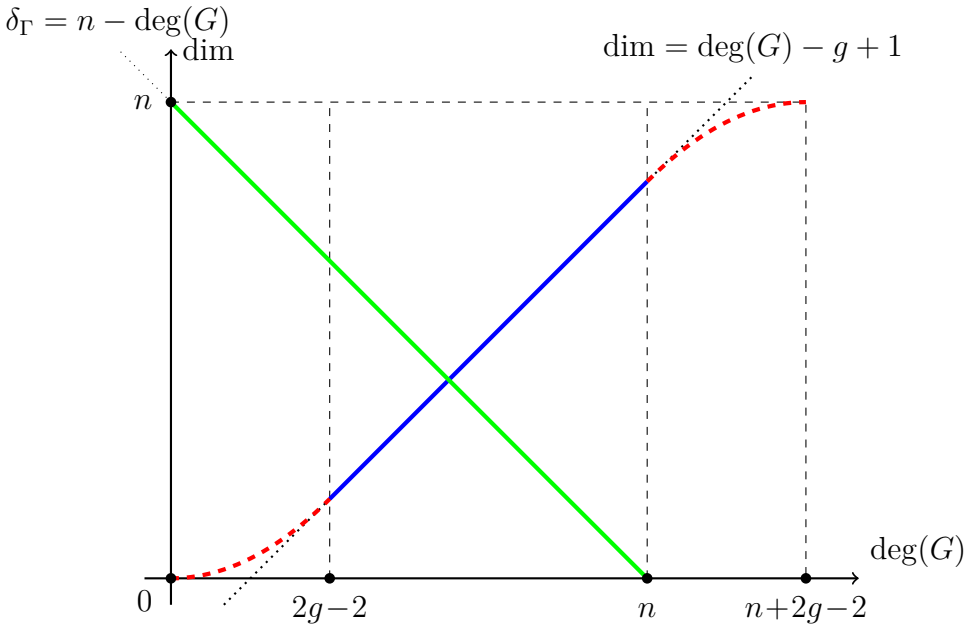
$$C_\Omega(D, G) = \left\{(res_{P_1}(dh), res_{P_2}(dh), ..., res_{P_n}(dh)) \mid dh \in \Omega_{\mathbb{F}_q(\mathcal{X})}(G - D)\right\}.$$

Theorem 3.3 ([Sti09])**.** *$C_\Omega(D, G)$ is an $[n, k', d']$ code with parameters:*

$$k' = i(G - D) - i(G) \quad \text{and} \quad d' \geq \deg G - (2g - 2).$$

if $\deg G > 2g - 2$, we have $k' = i(G - D) \geq n + g - 1 - \deg G$. If moreover $2g - 2 < \deg G < n$ then $k' = n + g - 1 - \deg G$.

Figure 3.1: Dimension and designed minimum distance of AG codes



Theorem 3.4 ([Sti09]). *The dual of the code $C_{\mathscr{L}}(D,G)$ is the code $C_\Omega(D,G)$*

$$C_\Omega(D,G) = C_{\mathscr{L}}(D,G)^\perp.$$

Furthermore, the differential code $C_\Omega(D,G)$ is equivalent with the functional code $C_{\mathscr{L}}(D, W+D-G)$. In particular, they have the same dimension and minimum distance, even though this equivalence does not preserve the all-important properties of the code. The formula $k = \ell(G) - \ell(G-D)$ also provides a lower bound $\delta_\Gamma = n - \deg(G)$ for its minimum distance. The integer δ_Γ is called the *Goppa designed minimum distance* of the AG code.

We illustrate the behavior of the dimension k of $C_{\mathscr{L}}(D,G)$ depending on the degree of the divisor G by Figure 3.1. In fact, Theorem 3.2 implies the exact value $k = \deg(G) - g + 1$ provided $2g-2 < \deg(G) < n$. Furthermore, if $\deg(G) > n + 2g - 2$, then $k = n$. In the intervals $[0, 2g-2]$, and $[n, n+2g-2]$, the dimension depends on the specific structure of the divisor G.

3.2.1 Genus zero AG codes

An algebraic curve of genus-zero over $\mathbb{F}_q$ is isomorphic to a projective line. The associated functions field can be identified with the field of rational functions in one indeterminate $\mathbb{F}_q(z)$. In coding theory, the class of AG codes on a genus zero curve is defined as genus-zero AG codes [Sti09]. They can be obtained by the general construction of both classes of codes. Genus zero functional codes over $\mathbb{F}_q$ are known as generalized Reed-Solomon codes. Moreover, differential codes on a genus zero curve are said to be geometric Goppa codes [HVLP98], since they can be presented by the restriction of the dual of generalized Reed-Solomon codes to the prime field:

$$C_\Omega(D,G)|_{\mathbb{F}_r} = C_{\mathscr{L}}(D,G)^\perp|_{\mathbb{F}_r} = T_{\mathbb{F}_q/\mathbb{F}_r}(C_{\mathscr{L}}(D,G))^\perp.$$

3.2.2 On the decoding of AG codes

There were beneficial papers on decoding algorithms for AG codes that started in the 1980s. The BerleKamp-Massey algorithm [Mas69] is an efficient decoding algorithm of Reed-Solomon codes that is known as the error-locator polynomial decodes up to half of their minimum distance. If the number of the known error positions is strictly less than the minimum distance, then we obtain the values of errors simply by solving linear equations involving syndromes. A generalization of this method was made by error-locator functions on curves. Thus, it is not extraordinary that Reed-Solomon and AG codes benefit from similar decoding algorithms. The work on the decoding of AG codes seems to begin in 1986 when Driencourt gave a first decoding algorithm for codes on elliptic curves of characteristic 2 [Dri85] correcting $\lfloor(\delta_\Gamma - 1)/2\rfloor$ errors. By generalizing the work of Arimoto and Peterson [Pet60] on employing a locator polynomial to decode Reed-Solomon codes, Justesen, Larsen, Jensen, Havemose, and Høhold published

[Jus+89] in 1989 a decoding algorithm for a larger class of AG codes, which can correct up to $\lfloor(\delta_\Gamma - g - 1)/2\rfloor$ errors, moreover in improved version [Jus+92] the error capability is increased to $\lfloor(\delta_\Gamma - g/2 - 1)/2\rfloor$. This method was generalized to arbitrary curves by Skorobogatov and Vladut [SV90], and independently by Krachkovskii [Kra88], then extended by Duursma [Duu93a; Duu93b] to correct $\lfloor(\delta_\Gamma - 1)/2\rfloor - \sigma$ errors, where σ is the Clifford defect of the curve [Duu93b]Definition 3.7 (is approximately $g/4$). In 1993, Feng and Rao [FR93] gave a majority voting scheme allowing a decoding up to $\lfloor(\delta_\Gamma - 1)/2\rfloor$ errors. Duursma generalized this result to all AG codes [Duu93c]. An efficient algorithm was described by Sakata, Justesen, Madelung, Jensen and Høhold in [Sak+95] using a multidimensional generalization of Massey-Berlekamp algorithm done by Sakata [Sak90]. Kirfel and Pellikaan [KP95] noticed that one could decode beyond $\lfloor(\delta_\Gamma - 1)/2\rfloor$ errors for 1–point AG codes by studying the Weierstrass semigroup. The reader can refer to [HP95; HVLP98; Pel93] for more details on decoding methods.

3.3 Hermitian codes

An important class of AG codes that have good properties is the class of Hermitian codes. This class is constructed by employing Hermitian curves over a finite field. The Hermitian curve $\mathscr{H}_q$ over $\mathbb{F}_{q^2}$ in affine coordinates has the form

$$\mathscr{H}_q : Y^q + Y = X^{q+1}.$$

Its rational points are points of the projective plane $PG(2, q^2)$, satisfying the homogenous equation $Y^q Z + Y Z^q = X^{q+1}$. It is easy to verify that $\mathscr{H}_q$ is nonsingular, then its genus is $g = q(q-1)/2$ by the genus formula. With respect to the line $Z = 0$ at infinity, $\mathscr{H}_q$ has one infinite point $P_\infty = (0 : 1 : 0)$ and q^3 affine rational points $P_1, \ldots, P_{q^3}$, which

make the class of Hermitian curves interesting since they attain the maximal number of rational points for the famous Hasse-Weil bound [Men+13]. As usual, we also look at the curve $\mathscr{H}_q$ as the smooth curve defined over the algebraic closure $\bar{\mathbb{F}}_{q^2}$. Then, there is a one-to-one correspondence between the points of $\mathscr{H}_q$ and the places of the function field $\bar{\mathbb{F}}_{q^2}(\mathscr{H}_q)$.

With a Hermitian code we mean a functional AG code of the form $C_{\mathscr{L}}(D, G)$, where the divisor D is defined as the sum $P_1 + \cdots + P_{q^3}$ of all affine rational points of $\mathscr{H}_q$. In our investigations, the divisor G can take two forms. In the *1-point case,* we set $G = sP_\infty$ with integer s. In the *degree 3 case,* we put $G = sP$, where P is a place of degree 3. Let P_1, P_2, P_3 be the extensions of P in the constant field extension of $\mathbb{F}_{q^2}(\mathscr{H}_q)$ of degree 3. Then P_1, P_2, P_3 are degree one places of $\mathbb{F}_{q^6}(\mathscr{H}_q)$ and, up to labeling the indices, $P_{j+1} = \mathrm{Frob}(P_j)$ where Frob is the q^2-th Frobenius map and the indices are taken modulo 3. Also, P may be identified with the $\mathbb{F}_{q^2}$-rational divisor $P_1 + P_2 + P_3$ of $\mathbb{F}_{q^6}(\mathscr{H}_q)$. Functional AG codes of the form $C_{\mathscr{L}}(D, sP_\infty)$ and $C_{\mathscr{L}}(D, sP)$ will be called 1-point Hermitian codes, and Hermitian codes over a degree 3 place, respectively. In the 1-point case, the basis of the Riemann-Roch space $\mathscr{L}(sP_\infty)$ can be given explicitly by [Ste12]:

$$\mathcal{M}(s) := \left\{ x^i y^j \mid 0 \le i \le q^2 - 1, 0 \le j \le q - 1, qi + (q+1)j \le s \right\}.$$

In the degree 3 case, the Riemann-Roch space

$$\mathscr{L}(sP) = \left\{ \frac{f}{(\ell_1 \ell_2 \ell_3)^u} \mid f \in \mathbb{F}_{q^2}[X, Y], \deg f \le 3u, v_{P_i}(f) \ge v \right\} \cup \{0\}.$$

can be computed, see [KN13]. In this formula, $\ell_i = 0$ is the equation of the tangent line of $\mathscr{H}_q$ at P_i, and $s = u(q+1) - v$, $0 \le v \le q$.

The group $\mathrm{Aut}(\mathscr{H}_q)$ of all automorphisms of $\mathscr{H}_q$ is defined over $\mathbb{F}_{q^2}$. It is a group of

projective linear transformations of $PG(2, q^2)$, isomorphic to the projective unitary group $PGU(3, q)$. Furthermore, $\mathrm{Aut}(\mathscr{H}_q)$ acts doubly transitively on the set $\{P_\infty, P_1, \ldots, P_{q^3}\}$ of $\mathbb{F}_{q^2}$-rational points. As it was pointed out in [KN13], the automorphism group of $\mathscr{H}_q$ acts transitively on the set of degree 3 places of $\mathbb{F}_{q^2}(\mathscr{H}_q)$, as well. Hence, the geometry of a degree 3 place is independent on the choice of P. However, the stabilizer G_P of P in $\mathrm{Aut}(\mathscr{H}_q)$ is not transitive on the set of $q^3 + 1$ rational points. In fact, G_P is a cyclic group of order $q^2 - q + 1$ and the number of G_P-orbits on the set of rational points is $q + 1$. (See [CKT99; KN13], where [CKT99, Section 4.2] holds for any characteristic.)

3.3.1 1–point Hermitian codes: parameters and dual codes

The dimension k of 1–point Hermitian codes $\mathcal{H}(q^2, s)$ is the dimension of $\mathscr{L}(sP_\infty)$, which can be determined from Riemann-Roch Theorem [Sti09]. $\mathcal{H}(q^2, s)$ has length $n = q^3$, if $2g - 2 < s < n$ then the dimension is $k = s - g + 1$ and the minimum distance is $d = q^3 - s$.

Theorem 3.5 (Dual codes [Men+13])**.** *For $s \geq 0$ define $\tilde{s} = q^3 + q^2 - q - 2 - s$. The codes $\mathcal{H}(q^2, s)$ and $\mathcal{H}(q^2, \tilde{s})$ are dual to each other.*

In particular, if q is even and $s = (q^3 + q^2 - q - 2)/2$, the code $\mathcal{H}(q^2, s)$ is self-dual.

Definition 3.3. *Let $\mathcal{H}(q^2, s)$ be a 1–point Hermitian code, the subfield subcode of $\mathcal{H}(q^2, s)$ is*

$$C_{q,r}(s) = \mathcal{H}(q^2, s)|_{\mathbb{F}_r}.$$

In [PJ14], the authors present an algorithm to compute $\dim C_{q,r}(s)$. Using this algorithm, the dimension of $C_{4,2}(s)$ is determined for each $s = 0, \ldots, 71$.

In [Vlu91, Proposition 3.2], the author shows

$$\dim T_{\mathbb{F}_{q^2}/\mathbb{F}_r}(\mathcal{H}(q^2, q)) = 2m + 1,$$

where $q = 2^m$. In our notation, this means

$$\dim C_{q,r}(q^3 + q^2 - 2q - 2) = q^3 - (2m + 1).$$

In particular, $\dim C_{4,2}(70) = 59$, which is confirmed by [PJ14, Table 2]. In the same table, we find $\dim C_{4,2}(s) = 1$ for $s = 0, \ldots, 31$ and $\dim C_{4,2}(32) = 5$. In the next section, we prove a formula which implies these dimensions.

3.3.2 On the true dimension of the subfield subcodes of 1–point Hermitian codes

The following is the main result of [EKN19]

Theorem 3.6. *Let $C_{q,r}(s)$ be a subfield subcode of the Hermitian code $\mathcal{H}(q^2, s)$, $q = r^m$ is a prime power. Then*

$$\dim C_{q,r}(s) = \begin{cases} 1 & \text{for } s < \frac{q^3}{r} \\ 2m + 1 & \text{for } s = \frac{q^3}{r} \end{cases}$$

Proof. Since the constant polynomials are in $\mathscr{L}(sP_\infty)$ for all $s \geq 0$, we have $\dim C_{q,r}(s) \geq 1$. We first show that $\dim C_{q,r}(s) = 1$ for $s < \frac{q^3}{r}$. Fix an integer $0 < s < \frac{q^3}{r}$ and take an arbitrary element $(c_1, \ldots, c_{q^3}) \in C_{q,r}(s)$. Then there is an element $f \in \mathscr{L}(sP_\infty)$ such that for all $i = 1, \ldots, q^3$, one has $c_i = f(P_i) \in \mathbb{F}_r$. There is an element $\gamma \in \mathbb{F}_r$ such that $c_i = \gamma$ for at least q^3/r indices i. In other words, $f - \gamma \in \mathscr{L}(sP_\infty)$ has at least q^3/r zeros on the Hermitian curve $\mathscr{H}_q$. (In fact, a nonzero element of $\mathscr{L}(G)$ cannot have more than $\deg G$ zeros on the curve.) Therefore, $f - \gamma$ must be the constant zero polynomial, and $c_i = \gamma$ for all i. In particular, $C_{q,r}(s)$ consists of the constant vectors.

Now, we suppose that $s = q^3/r$. Recall that

$$T_{\mathbb{F}_{q^2}/\mathbb{F}_r}(X) = X + X^r + \cdots + X^{r^{2m-1}}$$

is the trace polynomial of $\mathbb{F}_{q^2}$ over $\mathbb{F}_r$. We define the polynomial

$$f_{d,\alpha}(X) = d + T_{\mathbb{F}_{q^2}/\mathbb{F}_r}(\alpha X)$$

where $d \in \mathbb{F}_r$, $\alpha \in \mathbb{F}_{q^2}$. As a polynomial in one variable, $f_{d,\alpha}$ maps $\mathbb{F}_{q^2}$ to $\mathbb{F}_r$. If the point P_i is given with affine coordinates $P_i = (a_i, b_i)$, then $f_{d,\alpha}(P_i) = f_{d,\alpha}(a_i) \in \mathbb{F}_r$ for all $i = 1, ..., q^3$. In other words, the evaluation vector

$$\mathbf{c}_{d,\alpha} = (f_{d,\alpha}(P_1), \ldots, f_{d,\alpha}(P_{q^3})) \in \mathbb{F}_r^n.$$

We claim that $f_{d,\alpha}(x) \in \mathscr{L}\left(\frac{q^3}{r} P_\infty\right)$. In fact,

$$\rho(x^{r^k}) = qr^k,$$

which is at most $qr^{2m-1} = q^3/r$ for $k \leq 2m-1$. Hence, all monomials of $f_{d,\alpha}(x)$ are in $\mathscr{L}\left(\frac{q^3}{r} P_\infty\right)$, and the claim follows.

From the last two properties of $f_{d,\alpha}$ follows that the evaluation vector $\mathbf{c}_{d,\alpha} \in C_{q,r}(q^3/r)$. Since the map $(d, \alpha) \mapsto \mathbf{c}_{d,\alpha}$ is linear over $\mathbb{F}_r$, and injective, we have $\dim C_{q,r}(q^3/r) = 2m+1$.

In the last step we show that the elements $\mathbf{c}_{d,\alpha}$ exhaust the subfield subcode $C_{q,r}(q^3/r)$.

Take an element $g \in \mathscr{L}\left(\frac{q^3}{r} P_\infty\right)$ whose evaluation vector

$$(g(P_1), \ldots, g(P_{q^3})) \in \mathbb{F}_r^n.$$

We can reduce the high degree y-terms by the Hermitian equation $y^{q+1} = x + x^q$. Thus, we can write that g in this form:

$$g(x, y) = \sum_{j<q} a_{i,j} x^i y^j.$$

Moreover, since $\rho(x^i y^j) \equiv j \pmod{q}$, if $j \leq q-1$ then the value $\rho(x^i y^j)$ determines i and j uniquely. Therefore, each term of $g = \sum_{j \leq q-1} a_{i,j} x^i y^j$ has a different ρ–value. By definition we have

$$\rho(x^i y^j) = v_{P_\infty}(x^i y^j) = i\, v_{P_\infty}(x) + j\, v_{P_\infty}(y) = qi + (q+1)j.$$

The valuation of g at P_∞ is

$$v_{P_\infty}(g) = v_{P_\infty}\left(\sum a_{i,j} x^i y^j\right) = \max_{a_{i,j} \neq 0} \left(v_{P_\infty}(x^i y^j)\right).$$

Equality holds since the ρ–values are different. If $g \in \mathscr{L}\left(\left(\frac{q^3}{r} - 1\right) P_\infty\right)$ then $g = f_{d,0}$ for

some $d \in \mathbb{F}_r$ as seen above. Assume now

$$g \in \mathscr{L}\left(\frac{q^3}{r}P_\infty\right) \setminus \mathscr{L}\left(\left(\frac{q^3}{r}-1\right)P_\infty\right).$$

Then, $v_{P_\infty}(g) = q^3/r$ and g has a unique term $\beta x^{\frac{q^2}{r}}$ with ρ–value q^3/r, $\beta \in \mathbb{F}_{q^2}^*$. Define $\alpha \in \mathbb{F}_{q^2}$ by $\alpha^{r^{2m-1}} = \beta$. Then, $g - f_{0,\alpha} \in \mathscr{L}\left(\frac{q^3}{r}P_\infty\right)$ and $g - f_{0,\alpha}$ is again a constant $d \in \mathbb{F}_r$. This means $g = f_{d,\alpha}$, and the result follows. □

Using similar methods, we can show that for any $\alpha \in \mathbb{F}_{q^2}$,

$$T_{\mathbb{F}_{q^2}/\mathbb{F}_r}(\alpha y) \in \mathscr{L}\left(\frac{(q+1)q^2}{r}P_\infty\right).$$

Hence, $\dim C_{q,r}((q+1)q^2/r) \geq 4m+1$. By [PJ14, Table 2], we have equality for $q = 4$ and $r = 2$.

In Table 3.1 we present empirical results concerning the true dimension of the subfield subcodes of Hermitian codes for the parameters $q = 8$ and $r = 2$.

s	$\dim C_{8,2}(s)$	$\dim \mathcal{H}(64, s)$		s	$\dim C_{8,2}(s)$	$\dim \mathcal{H}(64, s)$
256	7	229		456	206	429
288	13	261		457	212	430
292	19	265		458	218	431
320	25	293		460	224	433
324	28	297		462	226	435
328	34	301		464	232	437
336	36	309		466	238	439
352	42	325		468	244	441
356	48	329		470	250	443
360	54	333		472	256	445
364	60	337		473	262	446
368	66	341		474	268	447
376	72	349		475	274	448
378	74	351		480	280	453
384	80	357		482	286	455
392	86	365		484	292	457
400	92	373		486	295	459
402	98	375		488	301	461
408	104	381		489	307	462
410	110	383		490	313	463
416	116	389		491	319	464
418	122	391		492	325	465
420	128	393		493	331	466
424	134	397		496	337	469
428	140	401		498	343	471
432	146	405		500	349	473
434	152	407		502	355	475
436	158	409		504	361	477
438	164	411		505	367	478
440	170	413		506	373	479
442	176	415		507	379	480
444	182	417		508	385	481
448	188	421		509	391	482
450	194	423		510	397	483
452	200	425		511	403	484

Table 3.1: Parameters of $C_{8,2}(s)$ for $s \in \{256, \ldots, 511\}$

Chapter 4

Estimating the dimension of Hermitian subfield subcodes

In this chapter, we study the possibility of the application of subfield subcodes of Hermitian codes in the McEliece scheme. More precisely, we do the first step by investigating the true dimension of these codes for a broad spectrum of parameters, for partial results, see [EKN19; PJ14]. Our main observation is that the true dimension of subfield subcodes of Hermitian codes can be estimated by the extreme value distribution function.

We established an approximating formula of the true dimension of the subfield subcodes of Hermitian codes. We conducted an experimental study to analyze the datasets of the true dimension of different subfield subcodes of Hermitian codes. This analysis helped us to derive new properties of their structure and led to an approach that might be useful for further research and applications. Before we tackle our contribution, we need to describe the set up of statistical formulas such as moment and expectation by mean of the extended rate function of the underlying classes of subfield subcodes of Hermitian codes.

4.1 Moments of the extended rate of subfield subcodes

In order to make our notation consistent, we make the following conventions. Let $\mathcal{X}$ be an algebraic curve over $\mathbb{F}_q$ and D, G divisors such that the AG code $C_L(D, G)$ is well defined. Assume that the objects δ and γ determine the curve $\mathcal{X}$ and the divisors D, G in a unique way. Let s be an integer and $\mathbb{F}_r$ be a subfield of $\mathbb{F}_q$. Then,

$$C^{\gamma}_{\delta,r}(s) = C_L(D, sG)|_{\mathbb{F}_r}$$

denotes the $\mathbb{F}_q/\mathbb{F}_r$ subfield subcode of the AG code $C_L(D, sG)$. The length of $C^{\gamma}_{\delta,r}(s)$ is $n = \deg(D)$.

For the integer s, let

$$R(s) = R^{\gamma}_{\delta,r}(s) = \frac{\dim_{\mathbb{F}_r} C^{\gamma}_{\delta,r}(s)}{n}$$

denote the rate of the subfield subcode $C^{\gamma}_{\delta,r}(s)$. We extend $R^{\gamma}_{\delta,r}$ to $\mathbb{R}$ in the usual way: $R^{\gamma}_{\delta,r}(x) = R^{\gamma}_{\delta,r}(\lfloor x \rfloor)$.

Lemma 4.1. *Let g be the genus of $\mathcal{X}$ and define*

$$\alpha = \left\lceil \frac{n + 2g - 2}{\deg(G)} \right\rceil.$$

Then $R(x)$ is a monotone increasing function, with

$$R(x) = \begin{cases} 0 & \text{for } x < 0, \\ 1 & \text{for } x \geq \alpha. \end{cases}$$

Proof. If $s \deg(G) > n + 2g - 2$, then $\deg(D + W - G) < 0$, and

$$C_{\Omega}(D, G) \cong C_L(D, D + W - G) = \{0\}.$$

Hence, if $s \geq \alpha$, then $C_L(D, sG) = \mathbb{F}_q^n$ and $C_L(D, sG)|_{\mathbb{F}_r} = \mathbb{F}_r^n$. □

The following observation has been made in Theorem 3.6 for the special case of a one point divisor of a Hermitian curve. (See also [EKN19, Theorem 5.1].)

Lemma 4.2. *For $0 \leq x < n/(r \deg(G))$, we have $R(x) = 1/n$.*

Proof. As the divisor sG is positive for $s > 0$, the constant vectors are in $C_L(D, sG)|_{\mathbb{F}_r}$ and $R(s) \geq 1/n$ holds. Assume $R(s) > 1/n$, that is, the subfield subcode contains a non constant element $\boldsymbol{v} = (f(P_1), \ldots, f(P_n))$ with $f \in \mathscr{L}(sG)$. Since f cannot have more than $\deg(sG)$ zeros, $\boldsymbol{v}$ cannot have the same entry more than $s \deg(G)$ times. This implies $r \deg(sG) \geq n$. □

Lemma 4.1 implies that we can consider $R(x)$ as the distribution function of some random variable ξ, cf. [Shi16, Definition 1, Section 2.3].

Lemma 4.3. *Let $R(x)$ be the extended rate function of a class of subfield subcodes $C_L(D, sG)|_{\mathbb{F}_r}$. Define the integer α as in Lemma 4.1. Let ξ be a random variable with distribution function $R(x)$. Then*

$$\mathsf{E}(\xi) = \sum_{s=0}^{\alpha} 1 - R(s), \qquad \mathsf{E}(\xi^2) = \sum_{s=0}^{\alpha} (2s+1)(1 - R(s)).$$

Proof. This follows from [Shi16, Section 2.6, Corollary 2]. □

Remark. Considered as a distribution function, $R^{\gamma}_{\delta,r}(s)$ has an expectation $\mathsf{E}^{\gamma}_{\delta,r}$, a variance $\mathsf{Var}^{\gamma}_{\delta,r}$ and a standard deviation $\mathsf{D}^{\gamma}_{\delta,r}$. These constants can be computed from the true dimensions of the subfield subcodes using Lemma 4.3 and the well known formulas of random variables.

4.2 Computed true dimensions of Hermitian subfield subcodes

Let q be a prime power. We say that the object $\delta = q$ determines the Hermitian curve $\mathscr{H}_q$ over $\mathbb{F}_{q^2}$, together with the divisor D which is the sum of affine rational points of $\mathscr{H}_q$. The objects $\gamma = 1$-pt or $\gamma =$ deg-3 determine the divisor G to be equal either to the

rational infinite place P_∞, or the degree 3 Hermitian place P, respectively. That being said, for any integer s and subfield $\mathbb{F}_r$ of $\mathbb{F}_{q^2}$, the Hermitian subfield subcodes

$$C_{q,r}^{\text{1-pt}}(s) = C_L(D, sP_\infty)|_{\mathbb{F}_r}, \qquad C_{q,r}^{\text{deg-3}}(s) = C_L(D, sP)|_{\mathbb{F}_r}$$

are well defined and consistent with the notation of section 4.1. In chapter 3, we denoted $C_{q,r}(s)$ by $C_{q,r}^{\text{1-pt}}(s)$. All these codes are $\mathbb{F}_r$-linear codes of length $n = q^3$.

Let $R_{q,r}^{\text{1-pt}}(s)$ and $R_{q,r}^{\text{deg-3}}(s)$ be the true rates of the codes $C_{q,r}^{\text{1-pt}}(s)$ and $C_{q,r}^{\text{deg-3}}(s)$. Using the GAP [Gap] package `HERmitian` [NEK19], we have been able to compute the true dimension values of the codes $C_{q,q}^{\text{1-pt}}(s)$, $C_{q,q}^{\text{deg-3}}(s)$ for

$$q \in \{2, 3, 4, 5, 7, 8, 9, 11, 13\}$$

and the binary codes $C_{q,2}^{\text{1-pt}}(s)$, $C_{q,2}^{\text{deg-3}}(s)$ for

$$q \in \{2, 4, 8, 16\}.$$

As given in Lemma 4.3, we computed the expectations $\mathsf{E}_{q,q}^{\text{1-pt}}$, $\mathsf{E}_{q,2}^{\text{1-pt}}$, $\mathsf{E}_{q,q}^{\text{deg-3}}$, $\mathsf{E}_{q,2}^{\text{deg-3}}$, the variances $\mathsf{Var}_{q,q}^{\text{1-pt}}$, $\mathsf{Var}_{q,2}^{\text{1-pt}}$, $\mathsf{Var}_{q,q}^{\text{deg-3}}$, $\mathsf{Var}_{q,2}^{\text{deg-3}}$, and the standard deviations $\mathsf{D}_{q,r}^{\mathbf{1\text{-}pt}}$, $\mathsf{D}_{q,2}^{\text{1-pt}}$, $\mathsf{D}_{q,q}^{\text{deg-3}}$, $\mathsf{D}_{q,2}^{\text{deg-3}}$ for these true rates. The numerical results are shown in Table 4.1 for $q = 3, 4, 5, 7, 8, 9, 11, 13$ and $r = q$, and in Table 4.2 for $q = 2, 4, 8, 16$ and $r = 2$. In Figure 4.1, we present the ratios $\mathsf{E}_{q,r}^{\gamma} \deg(G)/n$ and $\mathsf{D}_{q,r}^{\gamma} \deg(G)/n$, where $\gamma \in \{\text{1-pt, deg-3}\}$. While our data sets are small, these figures motivate the following open problem.

Problem 4.1. Are there constants $c_1, c_2 > 0$ such that

$$\mathsf{E}_{q,q}^{\text{1-pt}} \approx \mathsf{E}_{q,q}^{\text{deg-3}} \approx c_1 q^3/\deg(G), \qquad \mathsf{D}_{q,q}^{\text{1-pt}} \approx \mathsf{D}_{q,q}^{\text{deg-3}} \approx c_2 q^3/\deg(G),$$

where $a \approx b$ means $a/b \to 1$ with $q \to \infty$.

Remark. Our data suggests that for small q, the choice $c_1 = 0.75$ and $c_2 = 0.2$ is sound.

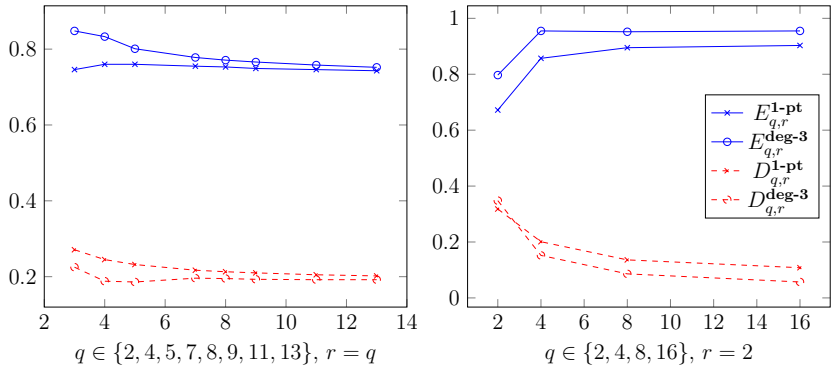
q	1-point codes		Codes over a degree 3 place	
	Expectation	Variance	Expectation	Variance
3	20.15	53.46	7.63	4.09
4	48.66	246.79	17.77	16.02
5	95.04	841.16	33.37	60.18
7	259.10	5 553.32	88.99	503.78
8	385.49	11 862.84	131.61	1 106.63
9	546.30	23 541.65	186.22	2 206.21
11	992.73	74 679.83	336.49	7 262.13
13	1 631.29	197 675.07	550.94	19 807.94

Table 4.1: Expectations and variances for Hermitian $\mathbb{F}_{q^2}/\mathbb{F}_q$ subfield subcodes

q	1-point codes		Codes over a degree 3 place	
	Expectation	Variance	Expectation	Variance
2	5.38	6.48	2.12	0.86
4	54.86	164.96	20.38	10.52
8	458.22	4 838.52	162.50	216.32
16	3 698.92	195 390.48	1 303.40	6 029.44

Table 4.2: Expectations and variances for Hermitian $\mathbb{F}_{q^2}/\mathbb{F}_2$ subfield subcodes

Figure 4.1: The ratios of expectations and standard deviations to $n/\deg(G)$



4.3 Distribution fitting

In general, no explicit formula is known for the true dimension of subfield subcodes of AG codes. We study the behavior of the subfield subcodes of Hermitian codes using distribution fitting methods.

As in the previous sections, we use the notation $\mathscr{H}_q$ for the Hermitian curve over $\mathbb{F}_{q^2}$, P_∞, P for the places of degree 1 and 3, D and $G \in \{P_\infty, P\}$ for the divisors, and $C^{\gamma}_{q,r}(s)$, $\gamma \in \{\text{1-pt, deg-3}\}$, for the $\mathbb{F}_{q^2}/\mathbb{F}_r$ subfield subcodes $C_L(D, sG)|_{\mathbb{F}_r}$. Then, with fixed q, r and $\gamma \in \{\text{1-pt, deg-3}\}$ the dimensions of the subfield subcodes are given by the extended rate functions

$$R^{\text{1-pt}}_{q,q}(x), \quad R^{\text{1-pt}}_{q,2}(x), \quad R^{\text{deg-3}}_{q,q}(x), \quad R^{\text{deg-3}}_{q,2}(x).$$

Our goal is to consider these functions as distribution functions and fit some well known probability distribution functions to our experimental rate function $R(x)$.

We obtain numerical results by using the distribution fitting methods offered by MATLAB's Statistics and Machine Learning Toolbox [TM19]. The technique MLE (Maximum Likelihood Estimation) is a method for estimating the parameters of a probability distribution from a data set. The method finds the parameter values maximizing the logarithm of the likelihood function [Eli93]. To compare different distributions for a given data set, one can use the log-likelihood values for a ranking. This is implemented MATLAB's `fitmethis` function [Cas20]. Notice that `fitmethis` also computes the AIC value for each distribution, which stands for the Akaike Information Criterion, that measures the quality of a model (distribution) versus the other models. It has the formula

$$AIC = 2l - 2\log(\hat{L})$$

where l is the number of parameters, and $\hat{L}$ is the maximum values of the likelihood

function. In the case of AIC, smaller values correspond to better-fitting distributions (see [KK08]).

In our comparisons, we restricted ourselves to parametric distributions having at most two parameters, that is, we used MATLAB's `fitmethis` to compare the log-likelihood values of the following distributions: normal, exponential, gamma, logistic, uniform, extreme value, Rayleigh, beta, Nakagami, Rician, inverse Gaussian, Birnbaum-Saunders, log-logistic, log-normal and Weibull. We can summarize the results as follows:

Proposition 4.4. *1. The best fitting distribution was the extreme value distribution for* $R_{q,q}^{1\text{-}pt}(x)$, $q \in \{4,5,7,8,9,11,13\}$, *for* $R_{q,q}^{deg\text{-}3}(x)$, $q \in \{7,8,9,11,13\}$, *and for* $R_{8,2}^{1\text{-}pt}(x)$, $R_{16,2}^{1\text{-}pt}(x)$, $R_{4,2}^{deg\text{-}3}(x)$, $R_{8,2}^{deg\text{-}3}(x)$, *and* $R_{16,2}^{deg\text{-}3}(x)$.

2. For the missing cases $R_{2,2}^{1\text{-}pt}(x)$, $R_{3,3}^{1\text{-}pt}(x)$, $R_{2,2}^{deg\text{-}3}(x)$, $R_{3,3}^{deg\text{-}3}(x)$, $R_{4,4}^{deg\text{-}3}(x)$, *and* $R_{5,5}^{deg\text{-}3}(x)$, *the best fitting distribution was the gamma distribution.*

3. The second best fitting distribution was the extreme value distribution for $R_{3,3}^{1\text{-}pt}(x)$, $R_{3,3}^{deg\text{-}3}(x)$, $R_{4,4}^{deg\text{-}3}(x)$, $R_{5,5}^{deg\text{-}3}(x)$.

Our results show that for $q \geq 3$, among the two-parameter distributions, the extreme value distribution function is a reasonable estimation of the rate function of subfield subcodes of Hermitian codes.

The extreme value distribution is also referred to as Gumbel or type 1 Fisher-Tippet distribution. In probability theory, these are the limiting distributions of the minimum of a large number of unbounded identically distributed random variables. The extreme value distribution function is

$$F(x;\alpha,\beta) = 1 - \exp\left(-\exp\left(\frac{x-\alpha}{\beta}\right)\right),$$

with location parameter $\alpha \in \mathbb{R}$ and a scale parameter $\beta > 0$. The mean μ and the variance σ^2 are

$$\mu = \alpha - \beta\gamma, \qquad \sigma^2 = \frac{\pi^2}{6}\beta^2,$$

where

$$\gamma = \int_1^\infty \left(-\frac{1}{x} + \frac{1}{\lfloor x \rfloor} \right) dx \approx 0.57721566490153$$

is the Euler-Mascheroni constant, see [KN00, Section 1.4]. With given empirical mean and variance of the data series, the parameters can be computed by

$$\alpha = \mu + \frac{\sqrt{6}\gamma}{\pi}\sigma, \qquad \beta = \frac{\sqrt{6}}{\pi}\sigma.$$

In figures 4.2 and 4.3, we visualized the fitting of the extreme value distribution function to our experimental results on the true dimension of subfield subcodes of Hermitian codes.

The occurrence of the extreme value distribution in the context of subfield subcodes of AG codes may be somewhat surprising, and we cannot give an understandable mathematical explanation for this. However, the rank of random matrices over finite fields is known to be related to the class of Gumbel type distributions; see Cooper's result [Coo00, Theorem 2] for the theoretical background. This theory has been applied to parameter estimates of random erasure codes by Studholme and Blake [SB10].

Figure 4.2: Estimating the extended rate function by extreme value distribution for subfield subcodes of 1-point Hermitian codes

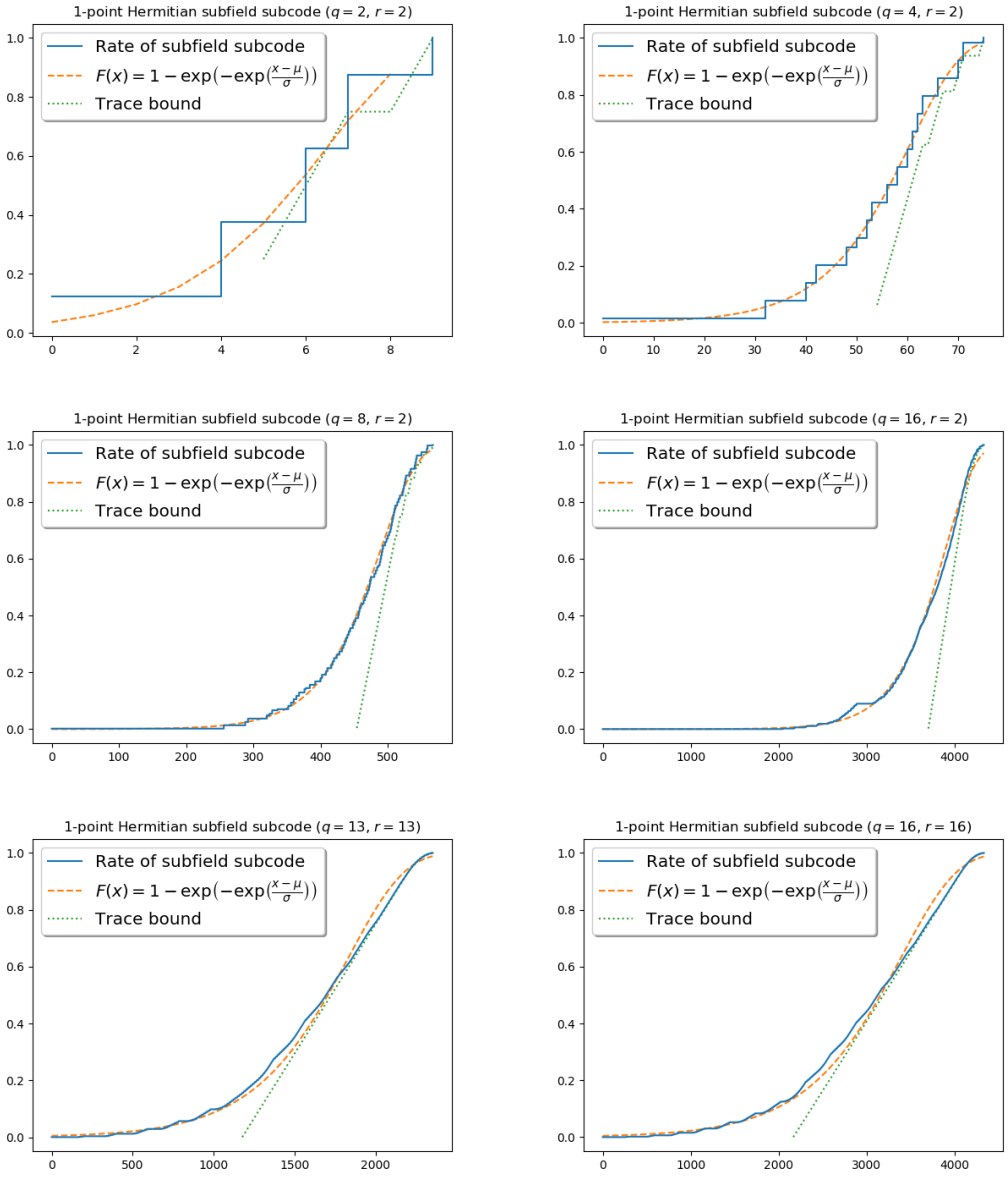
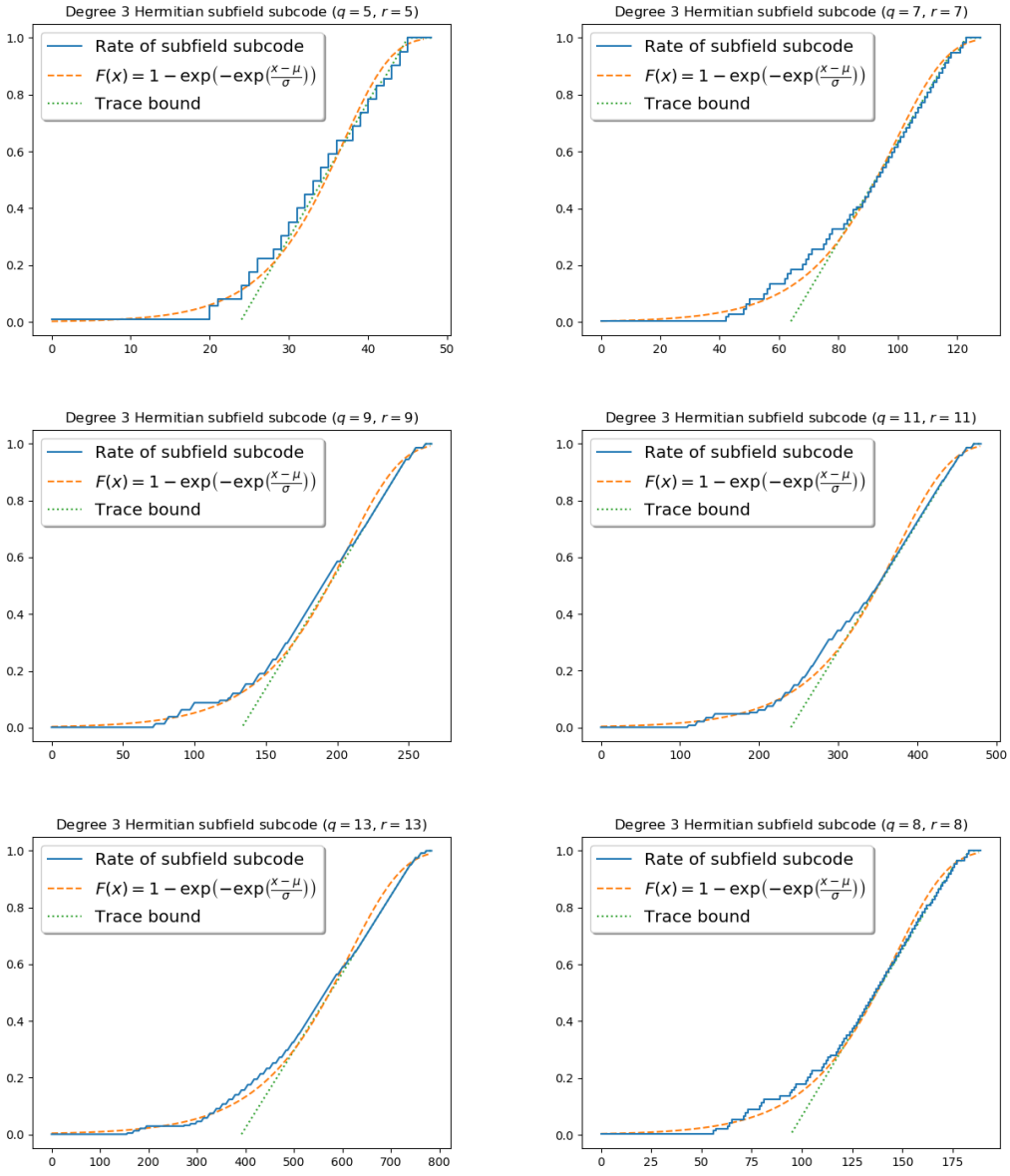


Figure 4.3: Estimating the extended rate function by extreme value distribution for subfield subcodes of degree 3 Hermitian codes by extreme value distribution



Chapter 5

McEliece cryptosystem: attacks and applications

This chapter provides the first step of our future work toward security analysis of McEliece cryptosystem based on Hermitian subfield subcodes. In the long term, we aim to make a comprehensive study in which we measure the McEliece cryptosystem security. Our attempt intends to improve the practicality of the underlying cryptosystem.

To assess the security of McEliece cryptosystem, we present some well-known attacks, for the reason that one of the security measurements of a cryptographic scheme is its resistance to standard cryptanalysis. The structure of this chapter is the following: we start with an overview of post-quantum cryptography [Nis; Aru+19]. In the second section, we describe two sorts of attacks: structural and decoding attacks. In the last section, we give an application of the subfield subcodes of Hermitian codes to cryptography. Mainly, we give a formula of the public key size in terms of the code rate using the result of section 4.

5.1 Post-quantum cryptography

In 1994 Shor [Sho94] introduced a quantum algorithm that is efficient in breaking cryptosystems which are believed to be secure for classical computers. Recently, the most frequent question is what sort of cryptosystems we can use in the presence of quantum computers. Once these latest will be available, we must have systems that are part of post-quantum techniques and which are known as post-quantum cryptography. It consists of different classes. Among them, we find:

- code-based cryptography,
- hash-based cryptography,
- lattice-based cryptography,
- multivariate-quadratic-equations cryptography.

Replacing these alternative systems will take time. Moreover, quantum-resistant cryptosystems should be in today's use to protect sensitive data. The construction of a secure cryptographic scheme must rely on a computationally hard problem. In classical cryptography, there are many schemes in which security is based on a difficult problem that a classical computer cannot solve, but a quantum computer can.

It seems not easy to design such schemes since its central aspect requires security for the systems to resist any attack. To attain this goal, the computational ability of an attacker should be taken into account. Also, the possession of a classical or quantum computer should be known. Thus, it is crucial to select secure classical cryptosystems if an adversary has a quantum computer. It is of great importance to think about secure cryptosystems that can be used for a classical computer and which will also remain secure

Cryptosystem	Broken by Quantum Algorithms
RSA public-key encryption	Broken
Diffie-Hellman key-exchange	Broken
Elliptic curve cryptography	Broken
Buchmann-Williams key-exchange	Broken
Algebraically Homomorphic	Broken
McEliece public-key encryption	Not broken yet
NTRU public key encryption	Not broken yet
Lattice-based public-key encryption	Not broken yet

Table 5.1: Current status of classical cryptosystems security in relation to quantum computers.

in the existence of quantum computers. Table 5.1 displays which classical cryptosystem will be secure in the quantum computing era. This also shows the significant attention given to McEliece's cryptosystem and its variants [BBD].

5.1.1 Code-based cryptography

Code-based cryptography is a set of cryptosystems in which the underlying trapdoor function is based on error-correcting codes. The first code-based cryptosystem was introduced by Robert J. McEliece in 1978. One must randomly select an error-correcting code to generate the private key that is the structure of the chosen code and the public key whose generator matrix has been randomly permuted. The plaintext is a codeword to which we add some errors in order to get ciphertext. Only the private key's possessor can decode the ciphertext to remove errors and recover the plaintext. It is required to adjust some parameters that concern its efficiency. Until now, there is no serious attack that threatens the security of the McEliece scheme even on quantum computers.

There were many attempts to design other cryptosystems with similar ideas. Among these cryptosystems, we mention: the Niederreiter scheme (which is a variant of McEliece cryptosystem with replacing the generator matrix by the parity check matrix of the

code [FD85]), the CFS signature scheme [CZ81], the identification schemes, and the cryptographic hash function [Ars+04].

In code-based cryptography, the practice is a trade-off between effectiveness and security at least for McEliece's cryptosystem. Also, it has many robust features such as security reduction is tight [BBD] and the encryption and decryption algorithms are very fast, i.e., they have low complexity.

Now, we take a look at the first code-based cryptosystem. We will briefly describe the McEliece cryptographic scheme since we aim at investigating the use of codes other than Goppa codes for this cryptosystem.

5.1.2 McEliece cryptosystem

McEliece introduced the first code-based public-key cryptosystem in 1978 where he employed error-correcting codes to generate the public and private key with security relying on two aspects: NP-completeness of decoding linear codes and distinguishing the chosen ones.

We consider a family of linear codes denoted by $\mathcal{F}$ with an efficient decoding algorithm. Let C be an element of $\mathcal{F}$ of length n and dimension k, with a decoding algorithm $\mathcal{A}_C$ of error capability t. The main idea behind McEliece's cryptosystem is that the sender applies the encoding process to the plaintext and then he adds some errors. Then, the receiver uses an encryption process that consists of the secret key, which is the decoding algorithm, to remove the errors and recover the plaintext.

For the generation of the keys, we consider the generator matrix G of C, a random $k \times k$ invertible matrix S and $n \times n$ permutation matrix P. Thus:

- **Public Key**: $(G' = SGP, t)$.

- **Secret Key**: G, S and P.

Let m be a plaintext of length k, and e a random error vector such that $wt(e) \leq t$.

- **Encryption**: $c = mG' = mSGP + e$.
- **Decryption**: to get back the original message m from c, we simply compute $(mSGP + e)P^{-1} = mSG + eP^{-1}$, then we decode to get mS. Thus $mSS^{-1} = m$.

The original McEliece cryptographic scheme is constructed on binary Goppa codes which are the subfield subcodes of generalized Reed-Solomon codes.

5.2 Attacks against code-based cryptography

In the literature, several attacks have been proposed against McEliece cryptosystem in general, and against McEliece systems that are based on AG codes in specific, see [BBC13]. Attacks can be divided into two classes: *structural* or key recovery attacks which aimed at recovering the secret code, and *decoding*, or message recovery attacks that seek to decrypt the transmitted ciphertext. The generic decoding attack against the McEliece scheme is the information set decoding (ISD) algorithm. The most recent and most effective structural attack against AG code-based McEliece systems is the Schur product distinguisher, which is given in [CMCP17], where the authors show that subfield subcodes of AG codes still resist. We focus on attacks based on Information Set Decoding (ISD) since they are useful for our case, and also, it is assumed to have the lowest complexity [Nie+12].

5.2.1 Structural attacks

In [CMCP17], the authors gave polynomial-time attacks against McEliece cryptosystem that relies either on AG codes or on their restriction to small fields. Their approach is inspired by the so-called *filtration attacks*. The idea of such an attack is based on the fact that AG codes can be distinguished from random ones by the mean of Schur product which is defined as follows:

Definition 5.1 ([CMCP17])**.** *The Schur product is the component wise product on $\mathbb{F}_q^n$: given two elements a and b in $\mathbb{F}_q^n$,*

$$a * b := (a_1 b_1, \cdots, a_n b_n).$$

For two codes $A, B \subseteq \mathbb{F}_q^n$ their Schur product is the code $A * B$ defined as

$$A * B := span_{\mathbb{F}_q}\{a * b | a \in A \text{ and } b \in B\}.$$

For $B = A$, then $A*A$ is denoted as $A^{(2)}$ and, we define $A^{(t)}$ by induction for any positive integer t.

This attack does not need to compute the structure of the curve and divisors that provide the public key codes. In their methods, the authors employed techniques based on filtration attacks to obtain an efficient decoding algorithm for AG codes which can be used as a public key.

For almost any AG code, the proof of the efficiency of this attack is held, except for AG codes on curves of large genus g with length satisfying $2g < n < 6g$, since there is no mathematical proof for such case. In [CMCP17, Section 7.3], the authors give a reason for which subfield subcodes are still resistant against this kind of attack. In the genus zero case, **GRS** subfield subcodes are still resistant to filtration attacks except for some

cases [COT16], then as well for the case of classical Goppa codes. Consequently, subfield subcodes of AG codes provide an interesting candidate for a secure McEliece scheme.

5.2.2 Decoding attacks (ISD)

In 1962 Prange introduced a generic decoding algorithm called *Information Set Decoding* which can solve the computational syndrome decoding (CSD) [Pet10], that consists of correcting t errors that occur in a codeword of a binary $[n, k]$ linear code. The decoding for linear codes is an NP-complete problem [BMT78], which is beneficial for code-based cryptosystems security [TS16]. The well-known algorithms that do not imply an explicit code structure are based on ISD. Briefly, the ISD procedure relies on selecting an information set which is a set of error-free coordinates in a codeword $c = x + e$ (the coordinate of c which are not different from the coordinates of x), in order to find an invertible sub-matrix formed by the corresponding columns of the generator matrix. Moreover, by solving linear equations, it is simple to recover the message m [TS16].

We denote by $\widehat{G} = (g_1^\top, ..., g_n^\top)$ the generator matrix in the systematic form. Let $m = (m_1, ..., m_k) \in \mathbb{F}_2^k$ be a plain message and y its encoded codeword in such a way that :

$$y = (y_1, ..., y_n) = m\widehat{G} + e \in \mathbb{F}_2^n$$

$e = (e_1, ..., e_n)$ is an added error. An attacker selects a random subset $\mathcal{I} \in \{1, ..., n\}$ of size k. $\mathcal{I}$ is an information set in case that $\widehat{G}_\mathcal{I}$ is invertible matrix. Then

$$y_\mathcal{I} = m\widehat{G}_\mathcal{I} + e_\mathcal{I}.$$

If $e_\mathcal{I}$ is non zero vector, then the procedure should be iterated with other information

sets. If not $e_{\mathcal{I}} = \mathbf{0}$, therefore it is easy to get m by computing:

$$y_{\mathcal{I}}\widehat{G}_{\mathcal{I}}^{-1} = m\widehat{G}_{\mathcal{I}}\widehat{G}_{\mathcal{I}}^{-1} + e_{\mathcal{I}}\widehat{G}_{\mathcal{I}}^{-1}$$

$\widehat{G}_{\mathcal{I}}^{-1}$ is the inverse of $\widehat{G}_{\mathcal{I}}$.

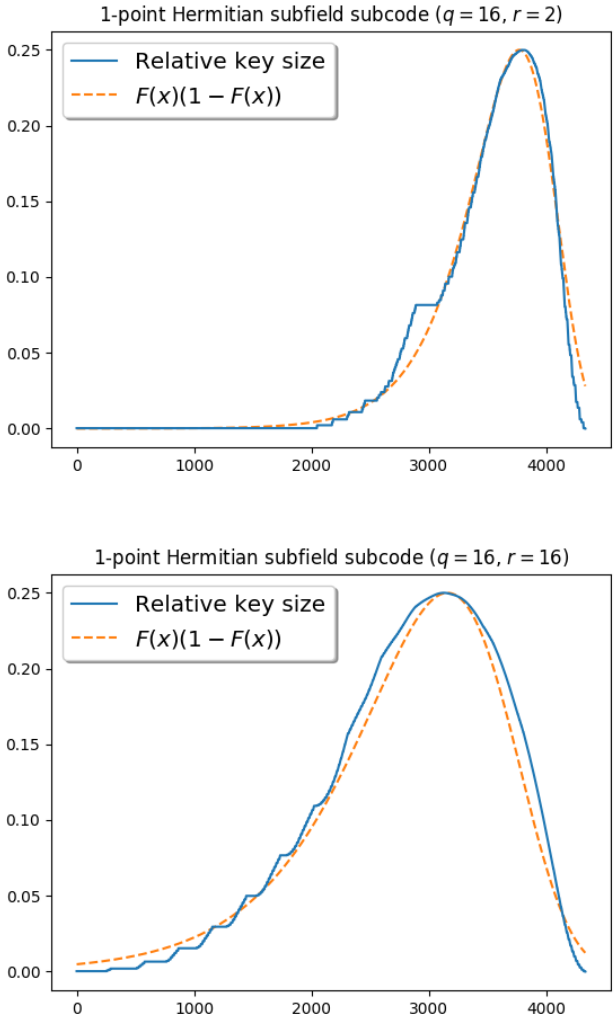
5.3 Selecting parameters to secure McEliece cryptosystem

We apply the result concerning the estimation of the true dimension of the subfield subcodes of Hermitian codes to estimate the key size of the McEliece cryptosystem. The largest (but not the only) part of the public key of the McEliece cryptosystem is the matrix G (see section 5.1.2). G is either the $n \times k$ generator matrix, or the $n \times (n-k)$ parity check matrix. In either case, G may be assumed to be in a systematic form, which means that the public key is given by $k(n-k)$ elements of $\mathbb{F}_r$. Hence, the key size is

$$\log_2(r)k(n-k).$$

In particular, for a fixed field $\mathbb{F}_r$ and length n, the key size is proportional to $R(1-R)$, see [Nie+12]. The true values of $R_{q,r}^{\gamma}(s)(1-R_{q,r}^{\gamma}(s))$ can be estimated by $F(x)(1-F(x))$, where $F(x)$ is the extreme value distribution function [EKN20], see Figure 5.1.

Figure 5.1: Estimating the key size $n^2R(1-R)$



Summary

Summary (in English)

The aim of this dissertation is the study of different classes of subfield subcodes of Hermitian codes. The results of our papers [EKN19], [EKN20] concern the structure and properties of these classes of error correcting codes. We also investigated the possible application of these codes in the McEliece cryptosystem. These are problems from the area of coding theory and code-based cryptography. Hermitian codes form a subclass of algebraic geometry (AG) codes, which can be defined by the function field of an algebraic curve over a finite field. The simplest case of an AG code is when the curve is a projective line, this gives the well-known class of Reed-Solomon codes. The subfield subcodes of the latter are the BCH and binary Goppa codes. Rather few precise results are known about the dimension of subfield subcodes of AG codes, and the known bounds typically hold only in a small parameter domain. In chapter 2 we present one of the few precise results, which are due to P. Véron ([Vlu90; Vlu91; VDV91]), and relate to certain classes of binary Goppa codes.

In chapter 3, we present the results of the paper [EKN19], where we determine the true dimension of subfield subcodes of 1-point Hermitian codes (Theorem 3.6) in the case when the parameter s satisfies $s \leq q^3/r$. Later, we extend the case of inequality for a broader class subfield subcodes (Lemmas 4.1 and 4.2).

In chapter 4, we rely on the paper [EKN20] which deals with the problem of approximating the true dimension of subfield subcodes of Hermitian codes by an explicit formula. We describe the statistical set up to tackle the experimental study to analyze the datasets of the true dimension of different subfield subcodes of Hermitian codes [NEK19]. Based on adjusting the distribution to the underlying datasets using the method `fitmethis` of MATLAB [TM19; Cas20], we found that the extreme value distribution is the most suitable one.

Recently, quantum computers and their algorithms are a real threat to the long-term confidentiality of our data. Only a few cryptosystems can resist this threat, one of these few is the McEliece cryptosystem, which is also the oldest and the best-known scheme. The classical version of it uses binary Goppa codes, and from a practical point of view, it suffers from the drawback of large key size. Chapter 5 is dedicated to these problems and the analysis of the parameters of subfield subcodes of Hermitian codes to give more precise estimates to the key size of the cryptosystem (Figure 5.1).

Magyar nyelvű összefoglaló (Summary in Hungarian)

Ezen disszertáció célja a Hermite-féle kódok résztest részkódjai különböző osztályainak vizsgálata. Az [EKN19] és [EKN20] cikkekben közölt eredményeinkben ezen kódosztályok struktúrájának és tulajdonságainak a vizsgálatát végeztük el. Vizsgáltuk továbbá ezen kódoknak a felhasználási lehetőségét a McEliece-féle titkosítási sémában. Ezek a problémák a kódoláselmélet és a kódalapú kriptográfia témakörébe esnek. A Hermite-kódok az algebrai-geometriai (AG) kódok osztályába tartoznak, amik egy véges test felett értelmezett algebrai görbe függvénytestének segítségévek definiálhatók. A legegyszerűbb esetben, amikor a görbe az egyenes, a jól ismert Reed-Solomon kódosztályt kapjuk. Ezek résztest részkódjai a BCH és a bináris Goppa kódok. Az AG kódok résztest részkódjainak

szélesebb osztályai esetén kevés pontos eredmény ismert a dimenzióra vonatkozóan, és a becslések is csak szűk paraméter tartományokra élesek. A 2. fejezetben ezen kevés pontos eredmények egyikét mutatjuk be, az eredmények P. Véron nevéhez fűzódnek ([Vlu90; Vlu91; VDV91], és bizonyos bináris Goppa kód osztályokra vonatkoznak.

A 3. fejezet az [EKN19] cikk eredményeit mutatja be. Ebben a Hermite-féle 1-pont kódok résztest részkódjainak pontos dimenzióját adjuk meg (Theorem 3.6) abban az esetben, ha az s paraméterre teljesül $s \leq q^3/r$. Az egyenlőtlenség esetét később kiterjesztjük szélesebb részkód osztályokra (Lemma 4.1 és 4.2).

A 4. fejezet az [EKN20] cikkre épül, amiben a Hermite-féle részkódok valódi dimenzióját közelítő explicit formulát kerestünk. A fejezetben bemutatjuk a statisztikai hátteret, valamint a kísérleti környezetet, amiben kis paraméterek esetén ($q \leq 16$) kiszámoljuk a valódi dimenziókat [NEK19]. Használva a MATLAB program statisztikai csomagját [TM19] és az ezt kiegészítő `fitmethis` [Cas20] függvényt, arra a következtetésre jutunk, hogy a kérdéses dimenzió az extrém érték eloszlás függvényével közelíthető (Proposition 4.4).

Napjainkban a kvantum algoritmusok és a kvantumszámítógépek és ezek speciális algoritmusai már komoly fenyegetést jelentenek az adatok hosszú távú titkosítására nézve. Ennek a veszélynek csak kevés kriptorendszer tud ellenállni, ezek kevesek közül a legrégibb és legismertebb a McEliece-féle séma. Ennek klasszikus verziója a bináris Goppa-kódokat használja, és gyakorlati szempontból nagy hátránya a nagy kulcsméret. Az 5. fejezetben ezt a kérdést jártuk körül, és vizsgáltuk általunk tekintett kódosztály paramétereit abból a célból, hogy pontosabb becsléseket tudjunk adni a rendszer kulcsméretére (Figure 5.1).

Appendix A

The GAP package HERmitian

`HERmitian` is a GAP package for computation in algebraic geometry codes theory developed by Gábor P. Nagy and Sabira El Khalfaoui. The package provides tools to work with Divisors and Riemann-Roch Spaces of Algebraic Function Fields of Hermitian Curves. This enables constructing different classes of AG codes over a Hermitian curve. HERmitian relies on several GAP packages, in particular on `OnAlgClosure` and `GZero`, both are implemented by Gábor P. Nagy.

A.1 Features

`HERmitian` provides the basic functionality for the following objects.

- **Hermitian curves:** indeterminate of a Hermitian curve, genus of a Hermitian curve affine and rational places of a Hermitian curve, random place, random place of a given degree.
- **Automorphisms of Hermitian curves:** Frobenius automorphism of a Hermitian curve, automorphism group of a Hermitian curve.

- **Hermitian divisors:** Hermitian place, Hermitian divisor, valuation, the zero divisor of a Hermitian curve, comparison of two Hermitian divisors, negative and positive part of a Hermitian divisor.

- **Hermitian Riemann-Roch spaces:** basis of Hermitian Riemann-Roch space, Hermitian AG functional and differential codes.

Among the crucial utilities of the `HERmitian` package is the possibility of computing Hermitian codes and their subfield subcodes.

A.2 Illustrations

The following example illustrates how to use `HERmitian` Package commands.

Let δ be the designed minimum distance of C, and define $t = [(\delta - 1 - g)/2]$. If there is a codeword $c \in C$ with $d(c, w) \leq t$ then c is returned. Otherwise, the output is `fail`.

The decoding algorithm is from [HP95]. The function `Hermitian_DECODER_DATA` precomputes two matrices which are stored as attributes of the AG code. The decoding consists of solving linear equations.

Example

```
gap> q:=4;
4
gap> # construct the curve and the divisors
gap> Y:=HermitianIndeterminates(GF(q^2),"Y1","Y2");
[ Y1, Y2 ]
gap> Hq:=Hermitian_Curve(Y[1]);
<Hermitian curve over GF(16) with indeterminates [ Y1, Y2 ]>
gap> P_infty:=Hermitian_Place(Hq,[infinity]);
<Hermitian place [ infinity ] over indeterminates [ Y1, Y2 ]>
gap>
gap> fr:=FrobeniusAutomorphismOfHermitian_Curve(Hq);
```

```
AC_FrobeniusAutomorphism(2^4)
gap> P4:=RandomPlaceOfGivenDegreeOfHermitian_Curve(Hq,5);;
gap> P4:=Sum(AC_FrobeniusAutomorphismOrbit(fr,P4));
<Hermitian divisor with support of length 5 over indeterminates [ Y1, Y2 ]>
gap> G:=5*P4+7*P_infty;
<Hermitian divisor with support of length 6 over indeterminates [ Y1, Y2 ]>
gap> Degree(G);
32
gap>
gap> len:=50;
50
gap> affpts:=AllRationalAffinePlacesOfHermitian_Curve(Hq);;
gap> D:=Sum(affpts{[1..len]});
<Hermitian divisor with support of length 50 over indeterminates [ Y1, Y2 ]>
gap>
gap> # construct the AG differential code
gap> Hermitian_DifferentialCode(G);
<[64,37] Hermitian AG-code over GF(2^4)>
gap> agcode:=Hermitian_DifferentialCode(G,D);
<[50,23] Hermitian AG-code over GF(2^4)>
gap> DesignedMinimumDistance(agcode);
22
gap> Length(agcode)-Degree(G)-1;
17
gap>
gap> # test codeword generation
gap> t:=Int((DesignedMinimumDistance(agcode)-1-Genus(G!.curve))/2);
7
gap> sent:=Random(agcode);;
gap> err:=RandomVectorOfGivenWeight(GF(q),Length(agcode),t);;
gap> received:=sent+err;;
gap>
gap> # decoding
gap> sent_decoded:=Hermitian_DecodeToCodeword(agcode,received);
<cvec over GF(2,4) of length 50>
gap> sent=sent_decoded;
true
```

Bibliography

[Ars+04] G. Ars et al. "Comparison between XL and Gröbner basis algorithms". In: *International Conference on the Theory and Application of Cryptology and Information Security*. Springer. 2004, pp. 338–353.

[Aru+19] F. Arute et al. "Quantum supremacy using a programmable superconducting processor". In: *Nature* 574.7779 (2019), pp. 505–510.

[BBC13] M. Baldi, M. Bianchi, and F. Chiaraluce. "Security and complexity of the McEliece cryptosystem based on quasi-cyclic low-density parity-check codes". In: *IET Information Security* 7.3 (2013), pp. 212–220.

[BMT78] E. R. Berlekamp, R. J. McEliece, and H. C. A. van Tilborg. "On the inherent intractability of certain coding problems". In: *IEEE Trans. Inform. Theory* IT-24.3 (1978), pp. 384–386.

[BBD] D. J. Bernstein, J. Buchmann, and E. Dahmen. *Post-Quantum Cryptography.–2009*.

[BS95] S. V. Bezzateev and N. A. Shekhunova. "Subclass of binary Goppa codes with minimal distance equal to the design distance". In: *IEEE Trans. Inform. Theory* 41.2 (1995), pp. 554–555.

[CZ81] D. G. Cantor and H. Zassenhaus. "A new algorithm for factoring polynomials over finite fields". In: *Mathematics of Computation* (1981), pp. 587–592.

[Cas20] F. de Castro. *fitmethis, Version 1.3.0.0*. MATLAB Central File Exchange. Jan. 2020.

[Coo00] C. Cooper. "On the distribution of rank of a random matrix over a finite field". In: *Proceedings of the Ninth International Conference "Random Structures and Algorithms" (Poznan, 1999)*. Vol. 17. 3-4. 2000, pp. 197–212.

[CKT99] A. Cossidente, G. Korchmáros, and F. Torres. "On curves covered by the Hermitian curve". In: *J. Algebra* 216.1 (1999), pp. 56–76.

[CMCP17] A. Couvreur, I. Márquez-Corbella, and R. Pellikaan. "Cryptanalysis of McEliece cryptosystem based on algebraic geometry codes and their subcodes". In: *IEEE Trans. Inform. Theory* 63.8 (2017), pp. 5404–5418.

[COT16] A. Couvreur, A. Otmani, and J.-P. Tillich. "Polynomial time attack on wild McEliece over quadratic extensions". In: *IEEE Transactions on Information Theory* 63.1 (2016), pp. 404–427.

[Del75] P. Delsarte. "On subfield subcodes of modified Reed-Solomon codes". In: *IEEE Trans. Information Theory* IT-21.5 (1975), pp. 575–576.

[Dri85] Y. Driencourt. "Some properties of elliptic codes over a field of characteristic 2". In: *International Conference on Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes*. Springer. 1985, pp. 185–193.

[Duu93a] I. M. Duursma. "Algebraic decoding using special divisors". In: *IEEE transactions on information theory* 39.2 (1993), pp. 694–698.

[Duu93b] I. M. Duursma. "Decoding–Codes from Curves and Cyclic Codes". In: (1993).

[Duu93c] I. M. Duursma. "Majority coset decoding". In: *IEEE transactions on information theory* 39.3 (1993), pp. 1067–1070.

[EKN19] S. El Khalfaoui and G. P. Nagy. "On the dimension of the subfield subcodes of 1-point Hermitian codes". In: *Advances in Mathematics of Communications* 0.0 (2019), p. 0. arXiv: `arxiv:1906.10444 [math.AG]`.

[EKN20] S. El Khalfaoui and G. P. Nagy. "Estimating The Dimension Of The Subfield Subcodes of Hermitian Codes". In: *Acta Cybernetica* (2020). To appear. arXiv: `arxiv:2004.05896 [math.AG]`.

[Eli93] S. R. Eliason. *Maximum likelihood estimation: Logic and practice*. 96. Sage, 1993.

[FD85] H. Fell and W. Diffie. "Analysis of a public key approach based on polynomial substitution". In: *Conference on the Theory and Application of Cryptographic Techniques*. Springer. 1985, pp. 340–349.

[FR93] G.-L. Feng and T. R. N. Rao. "Decoding algebraic-geometric codes up to the designed minimum distance". In: *IEEE Transactions on Information Theory* 39.1 (1993), pp. 37–45.

[Gap] *GAP – Groups, Algorithms, and Programming, Version 4.10.2*. The GAP Group, June 2019.

[Gop70] V. D. Goppa. "A new class of linear error-correcting codes". In: *Probl. Inf. Transm.* 6 (1970), pp. 300–304.

[HP95] T. Hoholdt and R. Pellikaan. "On the decoding of algebraic-geometric codes". In: *IEEE Transactions on Information Theory* 41.6 (1995), pp. 1589–1614.

[HVLP98] T. Høholdt, J. H. Van Lint, and R. Pellikaan. "Algebraic geometry codes". In: *Handbook of coding theory* 1.Part 1 (1998), pp. 871–961.

[Jus+89] J. Justesen et al. "Construction and decoding of a class of algebraic geometry codes". In: *IEEE Transactions on Information Theory* 35.4 (1989), pp. 811–821.

[Jus+92] J. Justesen et al. "Fast decoding of codes from algebraic plane curves". In: *IEEE Transactions on Information Theory* 38.1 (1992), pp. 111–119.

[KP95] C. Kirfel and R. Pellikaan. "The minimum distance of codes in an array coming from telescopic semigroups". In: *IEEE Transactions on information theory* 41.6 (1995), pp. 1720–1732.

[KK08] S. Konishi and G. Kitagawa. *Information criteria and statistical modeling*. Springer Science & Business Media, 2008.

[KN13] G. Korchmáros and G. P. Nagy. "Hermitian codes from higher degree places". In: *J. Pure Appl. Algebra* 217.12 (2013), pp. 2371–2381.

[KN00] S. Kotz and S. Nadarajah. *Extreme value distributions*. Theory and applications. Imperial College Press, London, 2000, pp. viii+187.

[Kra88] V. Y. Krachkovskii. "Decoding of codes on algebraic curves". In: *Conference Odessa*. 1988.

[LC01] S. Lin and D. J. Costello. *Error control coding*. Vol. 2. Prentice hall, 2001.

[MS77] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. I*. North-Holland Mathematical Library, Vol. 16. North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977, i–xv and 1–369.

[Mas69] J. Massey. "Shift-register synthesis and BCH decoding". In: *IEEE Transactions on Information Theory* 15.1 (1969), pp. 122–127.

[Men+13] A. J. Menezes et al. *Applications of finite fields*. Vol. 199. Springer Science & Business Media, 2013.

[Nag17] G. P. Nagy. *GZero, Divisors and Riemann-Roch spaces of Algebraic Function Fields of Genus Zero, Version 0.21*. GAP package. Oct. 2017.

[NEK19] G. P. Nagy and S. El Khalfaoui. *HERmitian, Computing with divisors, Riemann-Roch spaces and AG-odes of Hermitian curves, Version 0.1*. GAP package. Mar. 2019.

[Nie+12] R. Niebuhr et al. "Selecting parameters for secure McEliece-based cryptosystems". In: *International Journal of Information Security* 11.3 (June 2012), pp. 137–147.

[Pel93] R. Pellikaan. "On the efficient decoding of algebraic-geometric codes". In: *Eurocode'92*. Springer, 1993, pp. 231–253.

[Pet10] C. Peters. "Information-set decoding for linear codes over F q". In: *International Workshop on Post-Quantum Cryptography*. Springer. 2010, pp. 81–94.

[Pet60] W. Peterson. "Encoding and error-correction procedures for the Bose-Chaudhuri codes". In: *IRE Transactions on Information Theory* 6.4 (1960), pp. 459–470.

[PJ14] F. Piñero and H. Janwa. "On the subfield subcodes of Hermitian codes". In: *Designs, codes and cryptography* 70.1-2 (2014), pp. 157–173.

[Nis] *Post-Quantum Cryptography*. `http://csrc.nist.gov/projects/post-quantum-cryptography`. Updated: March 25, 2020.

[Ros+92] A. M. Roseiro et al. "The trace operator and redundancy of Goppa codes". In: *IEEE Trans. Inform. Theory* 38.3 (1992), pp. 1130–1133.

[Sak+95] S. Sakata et al. "Fast decoding of algebraic-geometric codes up to the designed minimum distance". In: *IEEE Transactions on Information Theory* 41.6 (1995), pp. 1672–1677.

[Sak90] S. Sakata. "Extension of the Berlekamp-Massey algorithm to N dimensions". In: *Information and Computation* 84.2 (1990), pp. 207–239.

[SMS97] T. Shibuya, R. Matsumoto, and K. Sakaniwa. "An improved bound for the dimension of subfield subcodes". In: *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences* 80.5 (1997), pp. 876–880.

[Shi16] A. N. Shiryaev. *Probability. 1.* 3rd ed. Vol. 95. Graduate Texts in Mathematics. Translated from the fourth (2007) Russian edition by R. P. Boas and D. M. Chibisov. Springer, New York, 2016, pp. xvii+486.

[Sho94] P. Shor. "Polynomial-time algorithm for prime factorization and discrete logarithms on a quantum computer: proc". In: *35th Annual Symp. on the Foundations of Computer Science*. Vol. 124. 1994.

[SV90] A. N. Skorobogatov and S. G. Vladut. "On the decoding of algebraic-geometric codes". In: *IEEE Transactions on Information Theory* 36.5 (1990), pp. 1051–1060.

[Ste12] S. A. Stepanov. *Codes on algebraic curves*. Springer Science & Business Media, 2012.

[Sti90] H. Stichtenoth. "On the dimension of subfield subcodes". In: *IEEE Transactions on Information Theory* 36.1 (1990), pp. 90–93.

[Sti09] H. Stichtenoth. *Algebraic function fields and codes*. Vol. 254. Springer Science & Business Media, 2009.

[SB10] C. Studholme and I. F. Blake. "Random matrices and codes for the erasure channel". In: *Algorithmica* 56.4 (2010), pp. 605–620.

[TM19] I. The MathWorks. *Statistics and Machine Learning Toolbox*. Natick, Massachusetts, United State, 2019.

[TS16] R. C. Torres and N. Sendrier. "Analysis of information set decoding for a sub-linear error weight". In: *Post-Quantum Cryptography*. Springer. 2016, pp. 144–161.

[VDV91] M. Van Der Vlugt. "A new upper bound for the dimension of trace codes". In: *Bulletin of the London Mathematical Society* 23.4 (1991), pp. 395–400.

[Vér98] P. Véron. "Goppa codes and trace operator". In: *IEEE Trans. Inform. Theory* 44.1 (1998), pp. 290–294.

[Véro01] P. Véron. "True dimension of some binary quadratic trace Goppa codes". In: *Des. Codes Cryptogr.* 24.1 (2001), pp. 81–97.

[Vér05] P. Véron. "Proof of conjectures on the true dimension of some binary Goppa codes". In: *Des. Codes Cryptogr.* 36.3 (2005), pp. 317–325.

[Vlu90] M. Van der Vlugt. "The true dimension of certain binary Goppa codes". In: *IEEE transactions on information theory* 36.2 (1990), pp. 397–398.

[Vlu91] M. van der Vlugt. “On the dimension of trace codes”. In: *IEEE Transactions on Information Theory* 37.1 (1991), pp. 196–199.